A Representation of exchangeable hierarchies by sampling from random real trees*

Chris Haulk[†] and Jim Pitman [‡]

January 31, 2011

Abstract

A *hierarchy* on a set S, also called a *total partition of* S, is a collection $\mathcal{H}$ of subsets of S such that $S \in \mathcal{H}$, each singleton subset of S belongs to $\mathcal{H}$, and if $A, B \in \mathcal{H}$ then $A \cap B$ equals either A or B or $\varnothing$. Every exchangeable random hierarchy of positive integers has the same distribution as a random hierarchy $\mathcal{H}$ associated as follows with a random real tree $\mathcal{T}$ equipped with root element 0 and a random probability distribution p on the Borel subsets of $\mathcal{T}$: given $(\mathcal{T}, p)$, let $t_1, t_2, \ldots$ be independent and identically distributed according to p, and let $\mathcal{H}$ comprise all singleton subsets of $\mathbb{N}$, and every subset of the form $\{j : t_j \in F_x\}$ as x ranges over $\mathcal{T}$, where F_x is the fringe subtree of $\mathcal{T}$ rooted at x. There is also the alternative characterization: every exchangeable random hierarchy of positive integers has the same distribution as a random hierarchy $\mathcal{H}$ derived as follows from a random hierarchy $\mathscr{H}$ on $[0, 1]$ and a family (U_j) of IID uniform [0,1] random variables independent of $\mathscr{H}$: let $\mathcal{H}$ comprise all sets of the form $\{j : U_j \in B\}$ as B ranges over the members of $\mathscr{H}$.

AMS 2010 subject classifications: 60E99, 60F99, 62E10, 62B05.
Keywords: exchangeable hierarchy, partition, random composition, fragmentation, continuum random tree, weighted real tree

1 Background

Definition 1.1. A *hierarchy* on a finite set S is a collection $\mathcal{H}$ of subsets of S such that

(a) if $A, B \in \mathcal{H}$ then $A \cap B$ equals either A or B or $\varnothing$, and

(b) $S \in \mathcal{H}$, $\{s\} \in \mathcal{H}$ for all $s \in S$, and $\varnothing \in \mathcal{H}$.

Hierarchies are known by several other names, including *total partitions* and *laminar families*. For brevity we use the term *hierarchy* throughout the paper. If $\mathcal{H}$ is a hierarchy on a finite set S and $S_0 \subseteq S$ then the *restriction of* $\mathcal{H}$ *to* S_0 is the hierarchy on S_0 defined as follows:

$$\mathcal{H}\Big|_{S_0} := \{H \cap S_0 : H \in \mathcal{H}\}. \tag{1}$$

*Research supported in part by N.S.F. grant DMS-0806118
[†]University of California at Berkeley; email haulk@stat.berkeley.edu
[‡]University of California at Berkeley; email pitman@stat.berkeley.edu

Definition 1.2. A *hierarchy on* $\mathbb{N}$ is a sequence $(\mathcal{H}_n, n \geq 1)$ where for each n, $\mathcal{H}_n$ is a hierarchy on $[n]$, and for every n, $\mathcal{H}_n = \mathcal{H}_{n+1}\Big|_{[n]}$.

Less formally, a hierarchy describes a scheme for recursively partitioning a set S into finer and finer subsets, down to singletons. If S is finite, this has an elementary meaning: S is partitioned into some set of blocks, then recursively: each non-singleton block that remains is partitioned into further blocks, until only singletons remain, and the hierarchy is the entire collection of sets that ever appear in this process. If S is infinite, matters can be more complex: a continuous recursive process of splitting may be involved, as in Bertoin's theory of self-similar or homogeneous fragmentation processes [14, 15] which have a natural regenerative structure. Alternatively, a hierarchy describes a process of coalescence, wherein the singleton subsets of S recursively coagulate to reconstitute the set S. We emphasize that *time* plays no role in our definition of a hierarchy: a hierarchy $\mathcal{H}$ encodes the *contents* of the blocks of some process of fragmentation (or coagulation), but does not include any additional information about the order in which these blocks appear in this fragmentation (or coagulation) process.

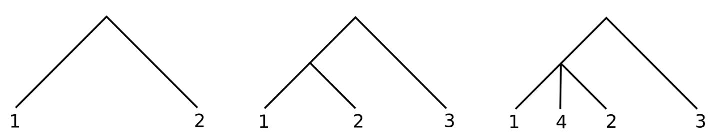

Figure 1: The tree on the right is the graph of $\mathcal{H} = \{\{1, 2, 4\}\} \cup \Xi([n])$. The other trees are the graphs of $\mathcal{H}\Big|_{[2]}$ and $\mathcal{H}\Big|_{[3]}$. The trivial hierarchy $\Xi([n])$ is defined at (3).

Hierarchies on $[n]$ are in bijective correspondence with certain trees. Explicitly, if $\mathtt{T}$ is a tree

- with n leaves, each labeled by a distinct element of $[n]$,
- and having a distinguished vertex called the root, which is not a leaf,
- with no internal vertices of degree two, except possibly the root
- and no edge lengths or planar embedding

then the map

$$\mathtt{T} \overset{g}{\mapsto} \{\{j \in [n] : \nu \text{ on path from leaf } j \text{ to root}\} : \nu \in V(\mathtt{T})\} \cup \{\varnothing\}$$

sends $\mathtt{T}$ to a hierarchy on $[n]$. Here, $V(\mathtt{T})$ denotes the set of vertices of $\mathtt{T}$ (including root and leaves). The map g is a bijection, and we we say that $\mathtt{T}$ is the *graph* of the hierarchy $g(\mathtt{T})$.

Random hierarchies of both finite and infinite sets arise naturally in a number of applications, including stochastic models for phylogenetic trees [57, 60, 8, 29, 29, 5, 4, 58, 67, 58], processes of fragmentation and coalescence [40, 3, 11, 12, 13, 16, 17, 18, 19, 20, 26, 30, 32, 35, 40, 42], and statistics and machine learning [1, 22, 44, 59, 21].

In these applications, the object of common interest is a rooted tree which describes evolutionary relationships (in the case of phylogentic trees) or the manner in which an object fragments into smaller pieces (in the case of models of fragmentation) or some notion of class membership (in the case of hierarchical clustering). Such trees sometimes have edges equipped with lengths that measure the time between speciations, or the amount of time between fragmentation events, or some measure of dissimilarity or distance between classes, but the hierarchies we consider correspond with trees of this type *without* edge lengths; see the remark in Section 5.

Permutations act on hierarchies by relabeling the contents of constituent sets: if $\mathcal{H}$ is a hierarchy on $[n]$ and σ a permutation of $[n]$, then

$$\sigma(\mathcal{H}) := \{\{\sigma(h) : h \in H\} : H \in \mathcal{H}\}.$$

An *exchangeable hierarchy on* $\mathbb{N}$ is a random hierarchy $(\mathcal{H}_n, n \geq 1)$ on $\mathbb{N}$ for which for every n and every permutation σ of n there is the distributional equality

$$\sigma(\mathcal{H}_n) \stackrel{d}{=} \mathcal{H}_n. \tag{2}$$

The purpose of this paper is to provide a de Finetti-type characterization of exchangeable random hierarchies on $\mathbb{N}$. Theorem 1 states that every exchangeable random hierarchy $\mathcal{H}$ on $\mathbb{N}$ is derived as if by sampling IID points $(t_j, j \geq 1)$ from a random measure μ supported by a random real tree $\mathcal{T}$: the blocks of $\mathcal{H}_n$ are the sets of the form $\{j \in [n] : t_j \in F_x\}$ as x ranges over $\mathcal{T}$, where F_x is the fringe subtree of $\mathcal{T}$ rooted at x. Real trees are tree-like metric spaces that are briefly discussed in Section 3.1; for a more complete treatment see [33] and references therein. Theorem 3 is an alternate characterization: every exchangeable hierarchy is derived as if from a sequence (U_j) of IID uniform[0,1] random variables and an independent random hierarchy $\mathscr{H}$ on [0,1]: the blocks of $\mathcal{H}_n$ are the sets of the form $\{j \in [n] : U_j \in B\}$ as B ranges over elements of $\mathscr{H}$. That $\mathscr{H}$ is a *random hierarchy on [0,1]* means simply that $\mathscr{H}$ is a random collection of subsets of $[0, 1]$ that satisfies (a) and (b) of Definition 1.1 with [0,1] in place of S. For some measure theory details concerning random hierarchies on [0,1] see the Remark at the end of Section 5.

As indicated in [10], an exchangeable hierarchy $(\mathcal{H}_n)$ of the set of positive integers $\mathbb{N}$ is generated by each of Bertoin's homogeneous fragmentation processes, and associated with each of Bertoin's homogeneous fragmentations there is a one-parameter family of self-similar fragmentations, each obtained from the homogeneous fragmentation by a suitable family of random time changes, and each generating the same random hierarchy $(\mathcal{H}_n)$ on $\mathbb{N}$. An attractive feature of the self-similar fragmentations of index $\alpha < 0$ is that each sample path of such a fragmentation is associated with a compact real tree [41]. The sample paths of Kingman's coalescent [57] can likewise be naturally identified with a compact real tree [32]. There has been considerable interest in describing real tree limits of discrete trees with edge-lengths [64, 43, 42, 68, 26], and Theorem 1 of this paper is in a similar vein.

This work forms part of a growing list of characterizations of infinite exchangeable combinatorial objects by de Finetti-type theorems. For example, Kingman characterized exchangeable partitions of $\mathbb{N}$ [56], Donelly and Joyce and Gnedin characterized composition structures [38, 28], Janson characterized exchangeable posets [50] and Hirth characterized exchangeable ordered trees [46], about which we will say a few words below.

Many related de Finetti-type theorems are known [6, 23, 24, 25, 34, 36, 37, 52, 61, 49], and there are excellent treatments in [7, 54] of related material. Such de Finetti-type results are often proved via reverse martingale convergence arguments, similar in spirit to the modern approach to de Finetti's theorem in [31, Chapter 4]. Alternate approaches use harmonic analysis [47, 48, 65, 66] or isometries of L^2 [9], or Choquet theory [45]. The results of this paper are proved using a third approach, the key idea of which is to encode an exchangeable hierarchy using a binary array, show that this array inherits exchangeability from the hierarchy, and apply well-known characterization theorems for arrays. A similar approach was first used by Aldous, who simplified of Kingman's proof characterizing of exchangeable partitions of $\mathbb{N}$ by encoding such partitions as exchangeable sequences of real random variables [7].

There are several papers on related topics. In [2, Theorem 3] it is shown that if $(\mathcal{R}(k), k \geq 1)$ is a consistent family of exchangeable trees with edge lengths that is *leaf-tight* then $(\mathcal{R}(k), k \geq 1)$ is derived as if by sampling from a random real tree. (Aldous also assumes that his trees are binary, but this assumption is not essential to his proof.) Since a hierarchy on $\mathbb{N}$ corresponds to a sequence of consistent trees *without* edge lengths, the main result of this paper can be seen as a variation on this result of Aldous, showing that leaf-tightness (and indeed any pre-defined notion of distance) is not needed to obtain a de Finetti type theorem for trees with exchangeable leaves.

In [19], it is shown that every *exchangeable $\mathcal{P}$-coalescent process* corresponds to a unique *flow of bridges*. An exchangeable $\mathcal{P}-$coalescent process is a Markov process $(\Pi_t, t \geq 0)$ whose state space $\mathcal{P}$ is the set of partitions of $\mathbb{N}$, for which Π_t is an exchangeable partition of $\mathbb{N}$ for every $t \geq 0$ whose increments are independent and stationary, if the notion of "increments" of a $\mathcal{P}$-valued function is properly understood. This provides a de Finetti-type characterization of exchangeable coalescents. One may "forget" time by setting $\mathcal{H} := \{B \subset \mathbb{N} : B \in \Pi_t \text{ for some } t > 0\} \cup \{\mathbb{N}\}$ and thereby obtain an exchangeable hierarchy $\mathcal{H}$ on $\mathbb{N}$ (the notation $B \in \Pi_t$ means that B is a block in the partition Π_t). The results of Bertoin and Le Gall in [19] therefore provide a de Finetti-type characterization of hierarchies that arise in this manner from exchangeable coalescents. Due to the stationary, independent increments property, this class of hierarchies is far from including every exchangeable hierarchy, so the present work may be seen as extending the results of Bertoin and Le Gall.

Haas and Miermont [41] provide a de Finetti-type representation of self-similar fragmentations of index $\alpha < 0$ that have no erosion or sudden loss of mass in terms of continuum trees $(\mathcal{T}, p)$ as follows: every such fragmentation $(F(t), t \geq 0)$ is derived as if from a continuum tree (T, p) by setting $F(t)$ equal to the decreasing sequence of masses of connected components of $\{t \in \mathcal{T} : \mathrm{ht}(v) > t\}$ where $\mathrm{ht}(t)$ denotes the distance from t to the root of $\mathcal{T}$. This is proved by introducing a family $(R(k), k \geq 1)$ of trees derived from an associated $\mathcal{P}$-fragmentation (Π_t) whose sequence of ranked limit frequencies equals $(F(t))$. Distances in these trees $R(k)$ are related to times between dislocations in $F(t)$, and by using self-similarity the *leaf-tight* criterion of [2] is checked. The existence of the representing tree $(\mathcal{T}, p)$ is then a consequence of the aforementioned theorem of Aldous. This provides a de Finetti-type theorem for self-similar fragmentations.

In [46], Hirth considers *exchangeable ordered trees*, which in our terms are exchangeable hierarchies $\mathcal{H}$ on $\mathbb{N}$ for which every element $B \in \mathcal{H}$ besides $\mathbb{N}$ there is an associated pair of nonnegative integer-valued times (N_B, M_B) which are the times at which B is "born" and the times at which B "dies." There is also a partial order on such blocks B

that is unimportant for our purposes. At the instant of its death, B gives birth to subsets born at that instant, whose union is B. Hirth provides a de Finetti-type characterization of exchangeable ordered trees using harmonic analysis techniques. Our hierarchies are more general than Hirth's trees, since there is no "discrete time" associated to the elements of a hierarchy. Our results may therefore be seen as an extension of Hirth's result using probabilistic techniques instead of harmonic analysis.

2 Results

This section provides some basic definitions and a statement of the main results of the paper.

For arbitrary sets S we define

$$\Xi(S) := \{S\} \cup \{\{s\} : s \in S\} \cup \{\varnothing\}. \tag{3}$$

We call $\Xi([n])$ the *trivial hierarchy on* $[n]$; it is the smallest hierarchy on $[n]$ and we will refer to it numerous times throughout the paper.

If $\mathcal{T}$ is a rooted real tree and $(t_n, n \in \mathbb{N})$ a deterministic or random sequence of points of $\mathcal{T}$, *the hierarchy derived from* $\mathcal{T}$ *and* $(t_n, n \in \mathbb{N})$ is the sequence $(\mathcal{H}_n, n \geq 1)$ defined by

$$\mathcal{H}_n := \{\{j \in [n] : t_j \in F_x(\mathcal{T})\} : x \in \mathcal{T}\} \cup \Xi([n]) \tag{4}$$

where $F_x(\mathcal{T})$ is the *fringe subtree of* $\mathcal{T}$ *rooted at* x,

$$F_x(\mathcal{T}) = \{y \in \mathcal{T} : x \text{ is in the geodesic path in } \mathcal{T} \text{ from } y \text{ to the root of } \mathcal{T}\}, \tag{5}$$

Real trees are tree-like metric spaces discussed in more detail in Section 3.1.

Recall that a random measure p is said to *direct* a family $(t_n, n \in \mathbb{N})$ if random elements if conditionally given p, $(t_n, n \in \mathbb{N})$ is an IID family with distribution p.

Theorem 1. *If $(\mathcal{H}_n, n \geq 1)$ is an exchangeable hierarchy on $\mathbb{N}$ then there is a $(\mathcal{H}_n)$-measurable triple $((\mathcal{H}'_n, n \geq 1), (\mathcal{T}, p), (t_1, t_2, \ldots))$, where $\mathcal{T}$ is a random real tree, p is a random probability measure with support contained in $\mathcal{T}$ almost surely, $(t_1, t_2, \ldots)$ is an exchangeable sequence directed by p, and $(\mathcal{H}'_n)$ is a hierarchy both equal in distribution to $(\mathcal{H}_n)$ and equal almost surely to the hierarchy derived from $\mathcal{T}$ and the samples $(t_1, t_2, \ldots)$.*

Theorem 1 is the main result of the paper, proved in Section 4, where we explicitly construct the pair $(\mathcal{T}, p)$. One of the issues in this construction is how *lengths* in $\mathcal{T}$ are defined. Our main device for defining lengths is the concept of most recent common ancestor.

Definition 2.1. If $\mathcal{H}_n$ is a hierarchy on $[n]$ and $i, j \in [n]$, then the *most recent common ancestor (MRCA) of* i *and* j, denoted $(i \wedge j)_n$, is the intersection of all elements of $\mathcal{H}_n$ that contain both i and j,

$$(i \wedge j)_n := \bigcap_{G \in \mathcal{H}_n : i, j \in G} G, \tag{6}$$

so, e.g., $(i \wedge i)_n = \{i\}$ if $i \geq n$. We adopt the convention that if one of i or j is not in $[n]$, $(i \wedge j)_n := \varnothing$. If $(\mathcal{H}_n, n \geq 1)$ is hierarchy on $\mathbb{N}$, then we denote by $(i \wedge j)$ the MRCA

of i and j in $(\mathcal{H}_n)$, which is the following set,

$$(i \wedge j) = \bigcup_{n \geq 1} (i \wedge j)_n \tag{7}$$

where $(i \wedge j)_n$ is the MRCA of i and j in $\mathcal{H}_n$. When no confusion can arise, we sometimes drop parentheses and subscripts from MRCAs to improve legibility. Also, when discussing more than one hierarchy, e.g. $(\mathcal{G}_n)$ and $(\mathcal{H}_n)$, we may write $(i \wedge j)_{\mathcal{G}_n}$ or $(i \wedge j)_{\mathcal{G}}$ to denote the MRCA of i and j in $\mathcal{G}_n$ or in $(\mathcal{G}_n)$.

To presage later developments, the family of indicators $\big(1(k \in (i \wedge j)), k \notin \{i, j\}\big)$ is exchangeable, so the limit

$$1 - \lim_{n\to\infty} \frac{1}{n} \#\{k \in [n] : k \in (i \wedge j)\}$$

exists almost surely. Also, the MRCA of i and j corresponds to a particular vertex in the graph of $\mathcal{T}_n$: the unique vertex found both in the path from root to leaf i and in the path from from root to leaf j that is at maximal graph distance from the root. This vertex has a counterpart ν, say, in the tree $\mathcal{T}$ of Theorem 1, and as will be made clear in the proof of that theorem, the distance from root to ν will the limit displayed above.

For comparison with Theorem 1, we state a version of Kingman's representation theorem for exchangeable partitions. Some preliminary definitions are necessary. Suppose that $\mathscr{P}$ is a fixed or random partition of $[0, 1]$ and that $(U_n, n \geq 1)$ is an IID sequence of uniform[0,1] random variables independent of $\mathscr{P}$. We say that a random partition Π of $\mathbb{N}$ is *derived as if by uniform sampling from* $\mathscr{P}$ if Π is equal in distribution to the partition of $\mathbb{N}$ which puts i and j in the same block if and only if U_i and U_j lie in the same block of $\mathscr{P}$. (We disregard for the moment the measure-theoretic details concerning *random* partitions of $[0, 1]$.)

A random partition Π of $\mathbb{N}$ is said to be *exchangeable* if the random array $(\mathbf{p}(i, j), i, j \in \mathbb{N})$ defined by

$$\mathbf{p}(i,j) = \begin{cases} 1 & i \text{ and } j \text{ are in same block of } \Pi \\ 0 & \text{else} \end{cases}$$

is exchangeable, meaning that for every $n \geq 1$ and every permutation σ of $[n]$

$$\Big(\mathbf{p}(\sigma(i), \sigma(j)), i, j \in [n]\Big) \stackrel{d}{=} \Big(\mathbf{p}(i, j), i, j \in [n]\Big). \tag{8}$$

The following is a weak version of Kingman's representation theorem for such partitions.

Theorem 2 ([56])**.** *If* Π *is an exchangeable partition of* $\mathbb{N}$ *then there is a* Π*-measurable random partition* $\mathscr{P}$ *of* $[0, 1]$ *for which* Π *is derived as if by uniform sampling from* $\mathscr{P}$*.*

A stronger version of this theorem is stated in Section 6.3. It is natural to ask whether Theorem 1 might be reformulated to resemble Theorem 2, and such a reformulation is indeed possible. Continuing to disregard measure-theoretic details, say that a random collection $\mathscr{H}$ of subsets of $[0, 1]$ is a *random hierarchy on [0,1]* if conditions (a) and (b) of Definition 1.1 hold with [0,1] in place of S. We say that $(\mathcal{H}_n)$ is *derived as if by uniform*

sampling from $\mathscr{H}$ if $(\mathcal{H}_n)$ is equal in distribution to the sequence of hierarchies $(\mathcal{H}'_n)$ defined by

$$\mathcal{H}'_n = \{\{j \in [n] : U_j \in B\} : B \in \mathscr{H}\},$$

where (U_n) is a sequence of IID uniform random variables independent of $\mathscr{H}$.

Theorem 3. *If $(\mathcal{H}_n)$ is an exchangeable hierarchy on $\mathbb{N}$, then there is an $(\mathcal{H}_n)$-measurable random hierarchy $\mathscr{H}$ on $[0,1]$ for which $(\mathcal{H}_n)$ is derived as if by uniform sampling from $\mathscr{H}$.*

Theorem 3 is proved in Section 5 as a Corollary of Theorem 1. The rest of the paper is organized as follows. Section 3 contains three subsections of definitions, well-known results, and elementary propositions needed for the proof of Theorem 1. Section 4 contains a proof of Theorem 1. Some complementary discussion and miscellaneous results may be found in Section 6.

3 Preliminaries

3.1 Real trees and hierarchies derived from real trees

Definition 3.1. A *segment* of a metric space X is the image of an isometry $\alpha : [a,b] \mapsto X$. The endpoints of the the segment are $\alpha(a)$ and $\alpha(b)$. A *real tree* is a metric space $(\mathcal{T}, d)$ for which

(a) for every pair x, y of distinct elements of $\mathcal{T}$ there is a unique segment with endpoints x and y, denoted $[[x,y]]$,

(b) if two segments of $\mathcal{T}$ intersect in a single point, and this point is an endpoint of both, then the union of these two segments is again a segment,

(c) If a segment contains distinct points u, v then it contains $[[u,v]]$,

(d) if the intersection of two segments contains at least two distinct points, then this intersection is a segment.

A real tree is *rooted* if there is a distinguished element of $\mathcal{T}$ called root. Every real tree we will discuss will assumed to be rooted, with root denoted 0. Furthermore, we define $[[x,x]] = \{x\}$.

In fact, parts (c) and (d) of Definition 3.1 follow from parts (a) and (b). For more regarding real trees see the excellent course notes [33]. The following example, however, provides sufficient background on real trees to understand the proof of Theorem 1.

Example 3.1 (Line-breaking and a random real tree, following Aldous)**.** Let ℓ_1 denote the Banach space of absolutely summable real sequences, and let $\mathbf{e}_i$ denote the ith element of the usual basis, so that $\mathbf{e}_1 = (1,0,0,\ldots)$, $\mathbf{e}_2 = (0,1,0,0,\ldots)$, and so on, and let (L_n) be a sequence of positive numbers. We define a family of real trees as follows: first, let $u_1 = (0,0,\ldots)$ and let

$$\mathcal{T}_1 = u_1 + \mathbf{e}_1[0, L_1] := \{(0,0,\ldots) + \mathbf{e}_1 x : 0 \le x \le L_1\}.$$

Next, select a point u_2 from $\mathcal{T}_1$ and let

$$\mathcal{T}_2 = \mathcal{T}_1 \cup u_2 + \mathbf{e}_2[0, L_2] := \mathcal{T}_1 \cup \{u_2 + \mathbf{e_2}x : 0 \le x \le L_2\}.$$

We continue recursively: supposing $\mathcal{T}_k$ has been defined, we select a point u_{k+1} from $\mathcal{T}_k$ and set

$$\mathcal{T}_{k+1} = \mathcal{T}_k \cup u_{k+1} + \mathbf{e}_{k+1}[0, L_{k+1}],$$

and let $\mathcal{T}$ be the closure of the union $\bigcup_{n\geq 1} \mathcal{T}_k$. The tree $\mathcal{T}_k$ is therefore built up by "gluing together" k line segments, and if we endow $\mathcal{T}$ with the ℓ_1 metric the geodesic paths in $\mathcal{T}_k$ flow along these line segments as one would expect.

The idea of using the natural basis of ℓ_1 in order to obtain a countable family of "orthogonal directions" in which to grow the new branch of $\mathcal{T}_k$, is due to Aldous [2].

To get a *random* real tree, simply randomize the construction above. For example, let (L_k) be the interarrival times of a Poisson process of on $\mathbb{R}_{\geq 0}$ of rate $t\,dt$, and for $k \geq 2$ select u_k according to normalized length measure on $\mathcal{T}_k$. The resulting tree is Aldous's Brownian continuum random tree.

We have defined in (4) the *hierarchy derived from $\mathcal{T}$ and a sequence $(t_n, n \in \mathbb{N})$ of points of $\mathcal{T}$*, but to make the definition precise we need to define the *fringe subtree of $\mathcal{T}$ rooted at a point* $x \in \mathcal{T}$, a concept used informally at (5).

Definition 3.2. If $\mathcal{T}$ is a real tree and x point of $\mathcal{T}$, then the *fringe subtree of $\mathcal{T}$ rooted at x* is the set

$$F_x(\mathcal{T}) := \{y \in \mathcal{T} : x \in [[0, y]]\}.$$

Proposition 4. *Let $\mathcal{T}$ be a real tree. Then for $x, y \in \mathcal{T}$, either $F_x(\mathcal{T}) \subset F_y(\mathcal{T})$, or $F_y(\mathcal{T}) \subset F_x(\mathcal{T})$, or $F_x(\mathcal{T}) = F_y(\mathcal{T})$.*

Proof. We claim that for all points $x, y, t \in \mathcal{T}$,

(i) if $x \in [[0, y]]$ and $y \in [[0, t]]$ then $x \in [[0, t]]$, and

(ii) if $x \notin [[0, y]]$ and $y \notin [[0, x]]$ then $F_x(\mathcal{T}) \cap F_y(\mathcal{T}) = \varnothing$.

If x, y, t are distinct non-root elements of $\mathcal{T}$ then (i) above follows from two applications of Part (c) of Definition 3.1. Likewise, if x, y, t are distinct non-root elements of $\mathcal{T}$ and $x \notin [[0, y]]$ and $y \notin [[0, x]]$, and $t \in F_x(\mathcal{T}) \cap F_y(\mathcal{T})$, then the segments $[[t, x]] \cup [[x, 0]]$ and $[[t, y]] \cup [[y, 0]]$ are distinct (y is not in the first, x is not in the second), and since these segments have the same endpoints we arrive at a contradiction with Part (a) of Definition 3.1, and (ii) follows. If x, y, t are not distinct or one if one or more of these is the root of $\mathcal{T}$, one may easily argue by cases. □

Corollary. *If $\mathcal{T}$ is a real tree and $(t_j, j \geq 1)$ a sequence of points of $\mathcal{T}$ then the sequence $(\mathcal{H}_n)$ defined by*

$$\mathcal{H}_n := \{\{j \in [n] : t_j \in F_x(\mathcal{T})\} : x \in \mathcal{T}\} \cup \Xi([n]),$$

is a hierarchy. Here, $\Xi([n])$ is the trivial hierarchy on $[n]$ defined at (3).

3.2 Random hierarchies: details

In this section we prove the following elementary proposition and show that hierarchies on $\mathbb{N}$ are in bijective correspondence with certain binary arrays.

Proposition 5. *1. If $n \geq 1$ and $\mathcal{H}_n$ is a hierarchy on $[n]$, then*

$$\mathcal{H}_n = \{(i \wedge j)_n : i, j \in [n]\} \cup \Xi([n])$$

where $\Xi([n])$ denotes the trivial hierarchy on $[n]$ and $(i \wedge j)_n$ the MRCA of i and j in $\mathcal{H}_n$.

2. If $(\mathcal{H}_n, n \geq 1)$ is a hierarchy on $\mathbb{N}$ then for every n

$$(i \wedge j) \cap [n] = (i \wedge j)_n,$$

where $(i \wedge j)$ and $(i \wedge j)_n$ denote the MRCAs of i and j in $(\mathcal{H}_n)$ and $\mathcal{H}_n$, respectively.

Proof. 1. Note that the subset of $\mathcal{H}_n$ consisting of sets that contain i is totally ordered by inclusion, by part (a) of Definition 1.1. The smallest member of this class that contains j is then $(i \wedge j)_n$. This shows that

$$\mathcal{H}_n \supseteq \{(i \wedge j)_n : i, j \in [n]\} \cup \Xi([n]).$$

To prove the reverse inclusion, fix $x \in \mathcal{H}_n$ and $i \in x$. The class $\{(i \wedge j)_n : j \in x\}$ is totally ordered by inclusion, with maximal element $(i \wedge j')_n$, say. Then for all $k \in x$, $k \in (i \wedge k)_n \subseteq (i \wedge j')_n$, so $x \subseteq (i \wedge j')_n$. On the other hand, $i, j' \in x$ and therefore $(i \wedge j')_n \subseteq x$. This proves the reverse inclusion.

2. By consistency of the sequence $(\mathcal{H}_n)$, for every $n \geq \max\{i, j\}$,

$$[n] \cap \bigcap_{G \in \mathcal{H}_{n+1} : \{i,j\} \subseteq G} G = \bigcap_{G \in \mathcal{H}_n : \{i,j\} \subseteq G} G.$$

It follows that $(i \wedge j)_n \subseteq (i \wedge j)_{n+1}$ for every positive n (recall $(i \wedge j)_n = \varnothing$ if $\max\{i, j\} \geq n$). The second assertion follows from this and the fact $(i \wedge j)_n \subseteq [n]$. $\square$

Proposition 5 shows that if $(\mathcal{H}_n)$ is a hierarchy on $\mathbb{N}$, then the class $\{(i \wedge j) : i, j \in \mathbb{N}\}$ contains complete information about $(\mathcal{H}_n)$, where $(i \wedge j)$ denotes the MRCA of i and j in $(\mathcal{H}_n)$. More explicitly, if the MRCA of i and j in $(\mathcal{H}_n)$ is known, then by restriction we obtain for every n the MRCA of i and j in $\mathcal{H}_n$, and $\mathcal{H}_n$ consists precisely of such MRCAs. The collection $\{(i \wedge j) : i, j \in \mathbb{N}\}$ can be conveniently encoded by the following array,

$$\mathtt{A}_{\mathcal{H}}(i, j, k) = \begin{cases} 1, & \text{if } k \in (i \wedge j) \\ 0, & \text{if } k \notin (i \wedge j) \end{cases} \qquad (i, j, k \in \mathbb{N}), \tag{9}$$

Such an array has two notable properties:

(a) For all triples $i, j, k \in \mathbb{N}$, $\mathtt{A}(i, j, k) = \mathtt{A}(j, i, k)$; also $\mathtt{A}(i, j, j) = 1$, and furthemore $\mathtt{A}(i, i, k) = 1$ if and only if $i = k$.

(b) For all pairs i, j and m, n of elements of $\mathbb{N}$, either the two sets

$$\{k \in S : \mathtt{A}(i,j,k) = 1\} \text{ and } \{k \in S : \mathtt{A}(m,n,k) = 1\}$$

are disjoint, or they are equal, or one of them contains the other.

Property (a) follows from symmetry of the roles of i and j in (6), and from the fact that $(i \wedge i) = \{i\}$. Property (b) follows from part (b) of Definition 1.1.

Proposition 6. *The correspondence (9) between hierarchies and binary arrays* $\mathtt{A} : \mathbb{N}^3 \mapsto \{0, 1\}$ *having properties (a) and (b) directly above, is bijective.*

The proof of this proposition is elementary and is therefore omitted.

3.3 Exchangeable Compositions

A *composition* of a set S is a partition of S together with a total order on blocks of this partition. Starting from such a pair one obtains a binary array $\mathtt{R}$ by setting $\mathtt{R}(i,j) = 1$ if either i and j are in the same block of the partition, or the block containing i precedes the block containing j, and otherwise setting $\mathtt{R}(i,j) = 0$, for all pairs $i, j \in S$. A binary array so derived necessarily has the following four properties, which hold for all $i, j, k \in S$.

(i) $\mathtt{R}(i,i) = 1$

(ii) if $\mathtt{R}(i,j) = 0$ then $\mathtt{R}(i,j) = 1$

(iii) if $\mathtt{R}(i,j) = 1$ and $\mathtt{R}(j,k) = 1$ then $\mathtt{R}(i,k) = 1$

(iv) if $\mathtt{R}(i,j) = 0$ and $\mathtt{R}(j,k) = 0$ then $\mathtt{R}(i,k) = 0$

Conversely, starting from a such an array $\mathtt{R}$ one may define an equivalence relation $\sim$ on S by

$$i \sim j \text{ if and only if } \mathtt{R}(i,j) = \mathtt{R}(j,i) = 1;$$

then the equivalence classes of $\sim$ form a partition of S, and we may totally order these classes by declaring that $[i]$ precedes $[j]$ if and only if $\mathtt{R}(i,j) = 1$, for all pairs i, j in S. This correspondence between a composition of a set S and a binary array $\mathtt{R}$ is obviously bijective. By (i)-(iv) above, the map

$$\mathtt{R} \mapsto \{(i,j) \in S^2 : R(i,j) = 1\}$$

sets up a bijective correspondence between such binary arrays $\mathtt{R}$ and binary relations on S that are reflexive, total, transitive, and whose complements are also transitive. Such relations need not be antisymmetric, and therefore need not be total orders, but every total order is such a relation. By abuse of notation we may use $i \ \mathtt{R} \ j$ and $\mathtt{R}(i,j) = 1$ interchangeably.

We will find it more convenient to work with arrays than with totally ordered set partitions or binary relations, so for our purposes, a *composition of a set S* will mean a binary array $\mathtt{R} : S \times S \mapsto \{0, 1\}$ for which properties (i)-(iv) above hold. If S is a finite or countably infinite set then an *exchangeable composition on S* is a random composition

$\mathtt{R}$ for which for every finite subset S_0 of S and every permutation σ of S_0, there is the distributional equality

$$\Big(\mathtt{R}(\sigma(i),\sigma(j)), i,j \in S_0\Big) = \Big(\mathtt{R}(i,j), i,j \in S_0\Big).$$

Theorem 7 below is a de Finetti-type characterization of exchangeable compositions, originally given in [38, Theorem 11] and [28, Theorem 5]. Before stating the theorem we must say a few words about *left-uniformization.*

By the *left-uniformization* F_* of a distribution F, we mean the image of F via the map $x \mapsto F_l(x)$ from $\mathbb{R}$ to $[0,1]$, where F_l denotes the left-continuous version of the distribution function of F. That is,

$$\begin{aligned} F_*[0,a] &= \mathbb{P}(F_l(X) \le a) \text{ for } X \text{ with distribution specified by} \\ & \quad \mathbb{P}(X \le x) = F(-\infty, x], \text{ and } F_l(x) = \lim_{w \uparrow x} F(-\infty, w] = F(-\infty, x). \end{aligned}$$

It is well-known that if F is a continuous distribution function, then F_* is the uniform distribution on $[0,1]$. More generally, if the discrete part of F has atoms of magnitude f_i and locations x_i, where $f_i \ge 0$ and $\sum_i f_i \le 1$, then F_* is characterized by the following three properties:

- (i) the distribution F_* has an atom of magnitude f_i at $u_i \in [0,1]$, where $u_i = F(-\infty, x_i)$, for each i;
- (ii) the distribution F_* places no mass on the interval $I_i := (u_i, u_i + f_i)$, for each i;
- (iii) the continous component of F_* is the restriction of Lebesgue measure on $[0,1]$ to the complement of $\cup_i I_i$.

We say that F is *left-uniformized* if $F_* = F$.

Theorem 7 ([38, Theorem 11] and [28, Theorem 5])**.** *If* $\mathtt{R}$ *is an exchangeable composition of* $\mathbb{N}$ *then the limit*

$$X_j = \lim_{m\to\infty} \frac{1}{m} \#\{n \in \{1,\dots,m\} : \mathtt{R}(j,n) = 0\} \tag{10}$$

exists almost surely for every $j \in \mathbb{N}$*. The family* $(X_j, j \in \mathbb{N})$ *so defined is exchangeable, and the directing measure of the family is left-uniformized with probability one. Furthermore, almost surely for all pairs* j,k*,* $\mathtt{R}(j,k) = 1$ *if and only if* $X_j \le X_k$*.*

Sketch of proof. For every $j \ge 1$, the family $(Y_n^j, n \ge 1)$ defined by

$$Y_n^j := \mathtt{R}(j,n') \quad n \in \mathbb{N}, \quad n' := \begin{cases} n & \text{if } n < j \\ n+1 & \text{if } n \ge j \end{cases}$$

is exchangeable, and $X_j = \lim_{m\to\infty} m^{-1} \sum_{k=1}^m Y_k^j$. The a.s. existence of the limit in (10) is therefore a consequence of de Finetti's theorem. Part of checking that the family (X_j) has the asserted properties involves showing that if $\mathtt{R}(j,i) = 0$ then $X_i < X_j$, and similar arguments using exchangeable sequences derived from $\mathtt{R}$ shows that this implication holds almost surely. The remainder of the argument is straightforward.

□

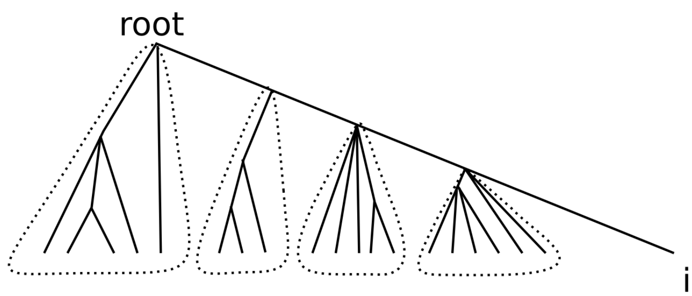

Figure 2: The ith spinal spinal composition associated to a hierarchy is the partition of leaves of the hierarchy into blocks according to attachment point on the spinal path from root to leaf i, together with the following ordering on these blocks: block s precedes block t if the attachment point for block s is nearer the root than the attachment point for block t.

3.4 Spinal Compositions

Definition 3.3. If $(\mathcal{H}_n, n \geq 1)$ is a hierarchy on $\mathbb{N}$ and i an element of S, the i^{th} *spinal composition of* $\mathbb{N} \setminus \{i\}$ is the binary array $\mathtt{R}_i$ defined by

$$\mathtt{R}_i(j,k) = \mathtt{A}_{\mathcal{H}}(i,j,k) \quad (j,k \in \mathbb{N} \setminus \{i\}) \tag{11}$$

where $\mathtt{A}$ is the binary array associated to $(\mathcal{H}_n)$ defined at (9).

The i^{th} spinal composition of $\mathbb{N} \setminus \{i\}$ associated to a hierarchy $(\mathcal{H}_n, n \geq 1)$ can be described less formally as follows in terms of the graph of $\mathcal{H}_n$ defined in Section 1.

> For $1 \leq i,j,k \leq n$, draw the path from root to leaf i in the graph of $\mathcal{H}_n$. Traverse the vertices of this path starting at root and moving towards i, and keep track of which vertices contain j and which contain k. If every vertex that contains j also contains k, then $R_i(j,k) = 1$, otherwise $R_i(j,k) = 0$.

See Figure 2 for a depiction of a spinal composition.

It is easily checked that $\mathtt{R}_i$ so defined is a composition of $\mathbb{N}\setminus\{i\}$. Furthermore, if $(\mathcal{H}_n)$ is an exchangeable hierarchy on $\mathbb{N}$ then $\mathtt{R}_i$ is an exchangeable composition of $\mathbb{N} \setminus \{i\}$. A version of Theorem 7 then holds, showing the existence of [0,1]-valued random variables

$$X^i_j := \begin{cases} \lim_{m\to\infty} \frac{1}{m} \#\{n \in [m] \setminus \{i\} : \mathtt{R}_i(j,n) = 0\} & j \in \mathbb{N} \setminus \{i\} \\ 1 & j = i \end{cases}. \tag{12}$$

The random variables $(X^i_j, j \in \mathbb{N} \setminus \{i\})$ are exchangeable and have a driving measure that is left-uniformized almost surely, and for $j,k \in \mathbb{N} \setminus \{i,j\}$, $\mathtt{R}_i(j,k) = 1(X^i_j \leq X^i_k)$ holds a.s. We call these variables (X^i_j) *spinal variables.*

Proposition 8. *Let $(\mathcal{H}_n)$ be an exchangeable hierarchy on $\mathbb{N}$, let $\mathtt{A}_{\mathcal{H}}$ and $(\mathtt{R}_i, i \in \mathbb{N})$ be the binary array associated to $(\mathcal{H}_n)$ as by (9) and the family of spinal compositions*

associated to $(\mathcal{H}_n)$, and for every $i \in \mathbb{N}$ let $(X^i_j, j \in \mathbb{N} \setminus \{i\})$ denote the family of spinal variables associated to $\mathtt{R}_i$ by (12). Then for all $i, j, k \in \mathbb{N}$ with $i \notin \{j, k\}$ there is the almost sure equality of events,

$$\{X^i_j \leq X^i_k\} = \{\mathtt{R}_i(j,k) = 1\} = \{\mathtt{A}_{\mathcal{H}}(i,j,k) = 1\} = \{(i \wedge k) \subseteq (i \wedge j)\}, \tag{13}$$

where for $i, j \in \mathbb{N}$, $(i \wedge j)$ denotes the MRCA of i and j in $(\mathcal{H}_n)$. Also,

$$X^i_j = \lim_{m \to \infty} \frac{1}{m} \#\{n \in \{1, \ldots, m\} : n \notin (i \wedge j)\} \tag{14}$$

holds with probability one for all distinct $i, j \in S$. Finally, with probability one, for distinct i, j, k, l in S,

(i) $X^i_j = X^j_i$ *if* $i \neq j$,

(ii) $(i \wedge j) = \{m \in S : X^i_m \geq X^i_j$ *or* $m = i\}$,

(iii) $X^i_k < X^j_k$ *implies* $X^i_k = X^i_j$.

Proof. The almost sure equalities in (13) are immediate consequences of definitions; (13) simply collects them in one place for easy reference. For (14) we note that for every distinct triple i, j, k of distinct elements of S, there is the almost sure equality of events

$$\{\mathtt{R}_i(j,n) = 0\} = \{n \notin (i \wedge j)\}. \tag{15}$$

which is immediate from (9) and (11). Now (14) follows from (15) and (12).

Assertion (i) follows from (14) and the fact that $(i \wedge j) = (j \wedge i)$. Assertion (ii) follows from (13).

For assertion (iii), suppose that $X^i_k < X^j_k$, then from (i) and (14) we have $(j \wedge k) \subsetneq (i \wedge k)$. We will show that $(i \wedge k) = (i \wedge j)$, by (14) this is enough for (iii). Already it is plain that $(i \wedge j) \subseteq (i \wedge k)$; it will suffice to show that $k \in (i \wedge j)$. We proceed by cases: since $(i \wedge k) \cap (i \wedge j) \neq \varnothing$, either $(i \wedge k) \subseteq (i \wedge j)$, in which case we are done, or $(i \wedge j) \subseteq (i \wedge k)$. So assume that $(i \wedge j) \subseteq (i \wedge k)$.

- If $(j \wedge k) \subseteq (i \wedge j)$, then $k \in (i \wedge j)$ and we are done.
- If $(i \wedge j) \subseteq (j \wedge k)$, then $(i \wedge k) \subseteq (j \wedge k)$, which is absurd, since $(j \wedge k) \subsetneq (i \wedge k)$.

Because $(j \wedge k) \cap (i \wedge j) \neq \varnothing$, one of the two bulleted cases above must obtain, and we conclude that $k \in (i \wedge j)$ as desired.

□

4 Proof of Theorem 1

Let $(\mathcal{H}'_n, n \geq 1)$ be an exchangeable hierarchy on $\mathbb{N}$. For reasons that will soon become clear, it will be much more convenient to work with a hierarchy on $\mathbb{Z}$ rather than on $\mathbb{N}$. Therefore fix an arbitrary bijection $b : \mathbb{N} \mapsto \mathbb{Z}$, and for every $n \geq 1$ set

$$\mathcal{H}_n := \{\{b(k) : k \in (i \wedge j)\} \cap [\pm n] : i, j \in \mathbb{N}\} \cup \Xi([\pm n]), \tag{16}$$

where $[\pm n] := \{-n, \dots, 0, \dots, n\}$ and Ξ is defined as at (3), and $(i \wedge j)$ is the MRCA of i and j in $(\mathcal{H}'_n)$. Then $\mathcal{H}_n$ is a hierarchy on $[\pm n]$ and $\mathcal{H}_{n+1}\Big|_{[\pm n]} = \mathcal{H}_n$ for every $n \geq 1$. We still need the notion of MRCA in $\mathcal{H}_n$ and in $(\mathcal{H}_n)$. Happily, Definition 2.1 makes sense in the present context with obvious minimal changes, e.g. reading $[\pm n]$ for $[n]$.

We will also need some auxiliary hierarchies, defined as follows.

Definition 4.1. For integers $i < 0$, $k < 0$, and $n \geq 1$, and Ξ defined at (3), and $(i \wedge l)$ the MRCA of i and l in $(\mathcal{H}_n)$ we set

$$\mathcal{H}_n^i := \{(i \wedge l) \cap [n] : l \in \mathbb{Z}\} \cup \Xi([n]), \tag{17}$$

$$\mathcal{G}_n^k := \{(i \wedge l) : i \in \{-1, \dots, k\}, l \in \mathbb{Z}\} \cup \Xi([n]) = \bigcup_{i=-1}^{k} \mathcal{H}_n^i. \tag{18}$$

$$\mathcal{G}_n := \{(i \wedge l) : i < 0, l \in \mathbb{Z}\} \cup \Xi([n]) = \bigcup_{i=-1}^{-\infty} \mathcal{H}_n^i. \tag{19}$$

It is easily checked that $\mathcal{H}_n^i$, $\mathcal{G}_n^k$, and $\mathcal{G}_n$ defined above are hierarchies on $[n]$. We now outline of the proof of Theorem 1.

(i) We define for every $k \leq 1$ a random tree $\mathcal{T}_k$ and a sequence $(t_j^k, j \geq 1)$ of random elements of $\mathcal{T}_k$. Both the tree and the samples are contained in ℓ_1, the Banach space of absolutely summable real sequences. We define another tree $\mathcal{T}$ and samples $(t_j, j \geq 1)$ by

$$\mathcal{T} := \mathrm{cl} \bigcup_{k \leq 1} \mathcal{T}_k, \qquad t_j := \lim_{k \to -\infty} t_j^k, \tag{20}$$

where cl denotes ℓ_1-closure, and the limits exist almost surely. Both $(t_j^k, j \geq 1)$ for $k \geq 1$ and $(t_j, j \geq 1)$ are exchangeable, and for the measure p of Theorem 1 we take the directing measure of the sequence $(t_1, t_2, \dots)$. For $k < -1$ we let p_k denote the directing measure of (t_j^k). These random measures (p_k) are not used in the proof of Theorem 1, but see Figure 3 for an image of how p_k and p_{k-1} are related.

(ii) We show that $\mathcal{G}_n$ is the hierarchy derived from $\mathcal{T}$ and the samples $(t_1, \dots, t_n)$, almost surely for all n. To do this, we first take the following intermediate step:

(ii a) We show that $\mathcal{G}_n^k$ is derived from $\mathcal{T}_k$ and the samples $(t_1^k, \dots, t_n^k)$, almost surely for all $n \geq 1$, $k \leq -1$. After taking this intermediate step, we establish the assertion of (ii) by taking a limit as $k \to -\infty$.

(iii) We show that the hierarchy $(\mathcal{G}_n, n \geq 1)$ is equal in distribution to the hierarchy $(\mathcal{H}'_n, n \geq 1)$ on $\mathbb{N}$ with which we started.

Steps (ii) and (iii), taken together, prove Theorem 1. The proof of Theorem 1 occupies the remainder of this section, and is broken into parts according the outline above.

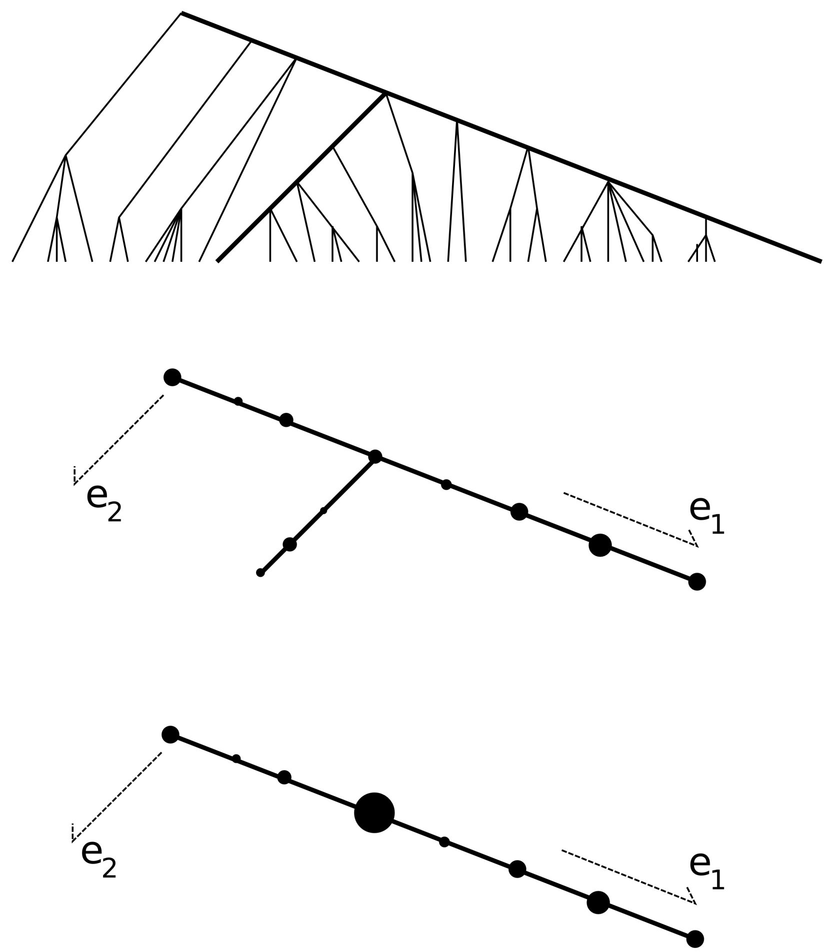

Figure 3: At top is shown the graph of $\mathcal{H}_n$ with leaf labels erased. The bold paths are the spinal paths to leaves -1 and -2, respectively. In the middle, $(\mathcal{T}_{-2}, p_{-2})$ is shown. The arrows indicate the ℓ_1 basis directions, and atoms of p_{-2} are represented by black circles or *beads* on $\mathcal{T}_2$, with circle size corresponding to atom size. At bottom is shown $(\mathcal{T}_{-1}, p_{-1})$. Note that (T_{-2}, p_{-2}) is derived from $(\mathcal{T}_{-1}, p_{-1})$ by "crushing" a bead on $\mathcal{T}_{-1}$ into fragments and stringing the crushed bead fragments out in the $\mathbf{e}_2$ direction.

4.1 Part (i)

Our main tool for constructing the real tree $\mathcal{T}$ and the samples $(t_1, t_2, \ldots)$ of Theorem 1 is the collection of $[0,1]$-valued spinal variables associated to spinal compositions, i.e. the family $(X_j^i, i, j \in \mathbb{Z}, i \neq j)$ defined by

$$X_j^i := \lim_{m\to\infty} \frac{1}{2m} \#\{n \in [\pm m] : n \notin (i \wedge j)\} \quad (i, j \in \mathbb{Z}, i \neq j) \tag{21}$$

where $(i \wedge j)$ denotes the MRCA of i and j in $(\mathcal{H}_n)$. We adopt the convention that $X_n^n \equiv 1$ for $n \in \mathbb{Z}$. Obviously Proposition 8 remains true in this context with minimal changes. It is worth emphasizing at this point that superscripts i and k on X_j^i's and t_j^i's, $\mathcal{G}_n^k$'s and $\mathcal{I}_n^k$'s (to be defined later) will be *negative*, and when taking limits we send k to $-\infty$ rather than ∞.

Definition 4.2. Let $(\mathbf{e}_j, j \geq 1)$ be the natural basis of ℓ_1, so that $\mathbf{e}_1 = (1, 0, 0, \ldots)$, $\mathbf{e}_2 = (0, 1, 0, 0, \ldots)$, etc., and for $m \geq 1$ let π_m denote the orthogonal projection onto $\operatorname{span}\{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$, so that $\pi_m((x_1, x_2, \ldots)) = (x_1, \ldots, x_k, 0, 0, \ldots)$, and $\pi_0((x_1, x_2, \ldots)) = (0, 0, \ldots)$.

Following Aldous [2], for $x \in \ell_1$ let $[[0, x]]_{sp}$ denote the path that proceeds from 0 to x along successive directions, for which $[[0, x]]_{sp}$ equals the closure of $[[0, x]]_{sp}^\circ$, where

$$[[0, x]]_{sp}^\circ := \bigcup_{m\geq 0} \{t\pi_m(x) + (1 - t)\pi_{m+1}(x) : 0 \leq t \leq 1\} \tag{22}$$

Observe that $[[0, x]]_{sp}$ differs from $[[0, x]]_{sp}^\circ$ only when $x = (x_1, x_2, \ldots)$ does not terminate in zeros, i.e when $x_j > 0$ for infinitely many j, and in this case the set difference $[[0, x]]_{sp} \setminus [[0, x]]_{sp}^\circ$ consists of the singleton $\{x\}$.

Definition 4.3. Let $(X_j^i, i, j \in \mathbb{N}, i \neq j)$ be the spinal variables defined in (21). For all $j \geq 1$, set $t_j^{-1} = \mathbf{e}_1 X_j^{-1}$ and for every $k \leq -2$ set

$$t_j^k := \mathbf{e}_1 X_j^{-1} + \sum_{l=2}^{k} \mathbf{e}_l \max\{0, X_j^l - \max\{X_j^{-1}, \ldots, X_j^{l-1}\}\} \quad (j \geq 1). \tag{23}$$

Define a family of trees $(\mathcal{T}_k, k \leq -1)$ using the samples $(t_j^k, j \geq 1, k \leq -1)$ as follows:

$$\mathcal{T}_k = \operatorname{cl} \bigcup_{j\geq 1} [[0, t_j^k]]_{sp},$$

where cl denotes closure in ℓ_1. For every $k \leq -1$ let d_k be the ℓ_1-metric on $\mathcal{T}_k$, and let $\mathcal{T}_k$ be rooted at $0 \in \ell_1$.

It is easily checked that for every $k \leq -1$, $(t_j^k, j \geq 1)$ is an exchangeable family. Observe that by definition, $||t_j^k|| = \max\{X_j^{-1}, \ldots, X_j^k\} \leq 1$.

Definition 4.3 of the samples (t_j^k) can be described as follows: once the samples (t_j^k) and tree $\mathcal{T}_k$ have been defined, to define (t_j^{k-1}) we select a subset of samples among those remaining and push these out in the $\mathbf{e}_{|k-1|}$-direction, orthogonal to $\mathcal{T}_k$ (this subset may possibly be empty). The next proposition shows that every one of these samples is selected from the *same spot* on $\mathcal{T}_k$; that is, $\mathcal{T}_{k-1}$ is derived by adding a *single* branch to $\mathcal{T}_k$ (or perhaps not adding a branch at all).

Proposition 9. *For every $k \leq -1$, the set $\{\pi_{|k|}(t_j^{k-1}) : t_j^{k-1} \neq t_j^k\}$ is either a singleton or the empty set.*

Proof. Suppose that $X_j^{k-1} > \max\{X_j^1, \ldots, X_j^k\}$. Then for every $i \in \{-1, \ldots, k\}$, $X_j^i < X_j^{k-1}$. Thus by part (iii) of Proposition 8, for every $i \in \{-1, \ldots, -k\}$, $X_j^i = X_{k-1}^i$. We have shown that

$$j \in \{\pi_k(t_j^{k-1}) : t_j^{k-1} \neq t_j^k\} \text{ implies } (X_j^1, \ldots, X_j^k) = (X_{k-1}^1, \ldots, X_{k-1}^k)$$

and we note that t_j^k is determined by $(X_j^1, \ldots, X_j^k)$ to conclude that $\{\pi_{|k|}(t_j^{k-1}) : t_j^{k-1} \neq t_j^k\}$ is a singleton. On the other hand, on the event that $X_j^{k-1} \leq \max\{X_j^1, \ldots, X_j^k\}$, for every $j \leq 0$ then $t_j^{k+1} = t_j^k$ for all j, and the set in question is empty. □

From the definition of $\mathcal{T}_k$ it can be seen that $\mathcal{T}_k$ is a real tree with probability one. It follows from Proposition 9 that $\mathcal{T}_k$ is furthermore a real tree derived by a line-breaking construction, like the tree in the Example in Section 3.1.

4.1.1 Examples

The following two examples are not part of the proof of Theorem 1 but together with Figure 3 they may help the reader visualize the construction of the tree $\mathcal{T}$.

Example 4.1. Let $(U_n, n \in \mathbb{Z})$ be a family of IID uniform[0,1] random variables, and let

$$\mathcal{H}_n := \{\{j \in [\pm n] : U_j \geq x\} : 0 \leq x \leq 1\} \cup \Xi([\pm n])$$

Following the construction above it can be seen that

$$\mathcal{T}_1 := \mathbf{e}_1[0, U_{-1}],$$

that p_{-1} is length measure on $\mathbf{e}_1[0, U_{-1})$, and that $\mathbf{e}_1 U_{-1}$ is an atom of p_{-1} of size $1 - U_{-1}$. Now let $k_1 = -1$ and define a sequence $(k_m, m \geq 1)$ recursively by $k_{m+1} := \max\{i < 0 : U_i > U_{k_m}\}$. Then $(\mathcal{T}_{-1}, p_{-1}) = \ldots = (\mathcal{T}_{k_2+1}, p_{k_2+1})$ and

$$\mathcal{T}_{k_2} := \mathcal{T}_1 \cup \mathbf{e}_1 U_{-1} + \mathbf{e}_{k_2}[0, U_{k_2} - U_{k_1}],$$

i.e., T_{k_2} is an isometric embedding of $[0, U_{k_2}]$ in ℓ_1, with a *kink* or bend at the image of U_{-1} in ℓ_1. The measure p_{k_2} is the sum of two measures: length measure on $\mathcal{T}_{k_2}$, and an atom of size $1 - U_{k_2}$ at the "end" $\mathbf{e}_{k_1} U_{k_1} + \mathbf{e}_{k_2} U_{k_2}$. In general, $\mathcal{T}_{k_m}$ is a an isometric embedding of $[0, U_{k_m}]$ into ℓ_1 with $|k_m| - 1$ kinks, and p_k is length measure on $\mathcal{T}_{k_m}$ plus an atom of size $1 - U_{k_m}$ at the end of T_{k_m}. The limit tree $\mathcal{T}$ is an isometric copy of $[0, 1]$, embedded in ℓ_1, and p is length measure on $\mathcal{T}$. The tree $\mathcal{T}$ has one *leaf* (nonroot element whose removal does not disconnect the space), and this leaf has p-measure 0.

Example 4.2. Let $\widehat{\mathcal{H}}$ denote the following collection of subsets of $[0, 1]$,

$$\widehat{\mathcal{H}} := \bigcup_{n \geq 1} \left\{ \left(\frac{j}{2^n}, \frac{j+1}{2^n} \right) : 0 \leq j \leq 2^n - 1 \right\} \cup \Xi([0, 1]).$$

Let $(U_n, n \in \mathbb{Z})$ be a family of IID uniform[0,1] random variables, and let

$$\mathcal{H}_n := \{\{j \in [\pm n] : U_j \in B\} : B \in \widehat{\mathcal{H}}\} \cup \Xi([\pm n])$$

Following the construction above it can be seen that

$$\mathcal{T}_1 := \mathbf{e}_1[0,1],$$

and that p_{-1} is purely atomic. The atoms can be described thusly: with $f_1 := 0$ and $f_n := \sum_{j=1}^{n-1} 2^{-n}$, the atoms of p_{-1} are at the locations $\{\mathbf{e}_1 f_n\}_{n\geq 1}$, and $p_{-1}(\{\mathbf{e}_n\}) = 2^{-n-1}$.

In fact, for every $k < 0$, the measure p_k is purely atomic. The atoms can be visualized as *beads* on the strings (segments) that constitute $\mathcal{T}_k$. To create the next tree $\mathcal{T}_{k-1}$, one of the atoms of p_k is selected with probability proportional to size and crushed into a sequence of smaller atoms, which are then strung out on the new string, respecting left-uniformization except at location of the crushed atom. More explicitly, suppose that $\mathcal{T}_k$ has been defined, and that the selected atom x has p_k-mass 2^{-m} for some m. It will follow from the construction that the distance from x to $0 \in \ell_1$ is $1 - 2^{-m+1}$, and that for some finite increasing sequence $1 \leq i_1 < i_2 \ldots < i_{j-1} < i_j = m$, x looks as follows,

$$x = (f_{i_1}, 0, f_{i_2} - f_{i-1}, 0, 0, 0, f_{i_3} - f_{i_2}, 0, 0, \ldots, f_{i_j} - f_{i_{j-1}}, 0, 0, 0, \ldots),$$

say; i.e. x is derived by thinning the vector $(f_{i_1}, f_{i_2} - f_{i-1}, f_{i_3} - f_{i_2}, \ldots, f_{i_j} - f_{i_{j-1}})$ with zeros. Suppose that $f_{i_j} - f_{i_{j-1}}$ is found in the lth coordinate of x; this indicates that the branch on which atom x is found was added at the lth step of the construction. Now, to create the next tree $\mathcal{T}_{k-1}$, set

$$\mathcal{T}_{k-1} = \mathcal{T}_k \cup x + \mathbf{e}_{|k|+1}[0, 2^{-m+1}],$$

ie. add a new branch at x in the $\mathbf{e}_{|k|+1}$-direction so that the total distance from root to tip of the new branch is 1. Note that for every point y in the new branch, there will be $|k| + 1 - l$ zeros between the penultimate and final nonzero entries of the y; this explains "zero-thinning".

The new measure p_{k-1} equals p_k on $\mathcal{T}_k \backslash \{x\}$. The atom x is crushed, and $p_{k-1}(x) = 0$; crushed bits of x are strung out on the new branch, so that p_{k-1} has atoms at the following locations,

$$(f_{i_1}, 0, f_{i_2} - f_{i-1}, 0, 0, 0, f_{i_3} - f_{i_2}, 0, 0, \ldots, f_{i_j} - f_{i_{j-1}}, 0, \ldots, 0, 2^n - 2^m, 0, 0, \ldots) \quad (n > m)$$

ie. at $x + \mathbf{e}_{|k|+1}(2^n - 2^m)$ for every $n \geq m$, and $p_{k-1}(x + \mathbf{e}_{|k|+1}(2^n - 2^m)) = 1 - 2^{-n-1}$.

The limit tree $\mathcal{T}$ has uncountably many leaves. The measure p is supported on these leaves, and on the set

$$\{x \in \ell_1 : \pi_j(x) \neq x \text{ for all } j \geq 1\},$$

because for every $j \geq 1$, all atoms on $\mathcal{T}_{-j}$ are eventually selected, crushed, and strung out, off of the set $\{x \in \ell_1 : \pi_j(x) = x\}$. It can also be seen that p is diffuse (i.e. nonatomic), because every atom is eventually crushed into smaller atoms, so no atom of positive mass can remain in the limit.

4.2 Part (ii)a

For $n \geq 1$ and $k \leq -1$, let $\mathcal{I}_n^k$ denote the hierarchy derived from $\mathcal{T}_k$ and the samples $(t_1^k, \ldots, t_n^k)$, that is,

$$\mathcal{I}_n^k := \{\{j \in [n] : x \in [[0, t_j^k]]_{sp}\} : x \in \mathcal{T}_k\} \cup \Xi([n]).$$

Proposition 10. *For all positive integers n and k, $\mathcal{G}_n^k = \mathcal{I}_n^k$ almost surely.*

Several intermediate results are needed to prove Proposition 10.

Lemma 11. *Let $\mathcal{H}$ be a hierarchy on a finite set S, and suppose that $i \in B \in \mathcal{H}$. Then almost surely there is $j \in S$ such that $B = (i \wedge j)$, where $(i \wedge j)$ denotes the MRCA of i and j in $\mathcal{H}$. As a corollary, if every element of $\mathcal{H}$ contains i, then*

$$\{(j \wedge l) : j, l \in S, j \neq l\} = \{(j \wedge s) : j \neq s\} \quad a.s. \tag{24}$$

Proof. Fix $B \in \mathcal{H}$. By the argument for Proposition 5,

$$B = \bigcup_{j \in B} (s \wedge j).$$

Since the members of $\{s \wedge j : j \in B\}$ all have the point s in common, by part (a) of Definition 1.1 they are totally ordered by inclusion. Therefore there is some maximal element j' of B for which $B = (s \wedge j')$. □

Lemma 12. *With $\mathcal{H}_n^i$ as in Definition 4.1,*

$$\{B \cap [n] : i \in B \in \mathcal{H}_n\} = \mathcal{H}_n^i = \{\{j \in [n] : X_j^i \geq x\} : 0 \leq x \leq 1\}$$

holds almost surely for all $n \geq 1$ and $i < 0$.

Proof. This follows from Lemma 11, Proposition 5, and part (ii) of Proposition 8. □

Lemma 13. *For all $k < 0$, and $x \in \mathcal{T}_k$,*

$$\{j > 0 : t_j^{k-1} \in F_x(\mathcal{T}_{k-1})\} = \{j \geq 0 : t_j^k \in F_x(\mathcal{T}_k)\}. \tag{25}$$

holds almost surely.

Proof. If t_j^{k-1} is in $F_x(\mathcal{T}_{k-1})$ for $x \in \mathcal{T}_k$ then $x \in [[0, t_j^{k-1}]]_{sp}$, so $x = \pi_{|k|}(x) \in \pi_{|k|}\Big([[0, t_j^{k-1}]]_{sp}\Big) = [[0, t_j^k]]_{sp}$, so $t_j^k \in F_x(\mathcal{T}_k)$. On the other hand, $[[0, t_j^k]]_{sp} \subseteq [[0, t_j^{k-1}]]_{sp}$ so $t_j^k \in F_x(\mathcal{T}_k)$ implies $t_j^{k-1} \in F_x(\mathcal{T}_{k-1})$. □

Proof of Proposition 10. Since $t_j^1 := \mathbf{e}_1 X_j^1$ for $j \geq 1$, for $k = -1$ the assertion is covered by Lemma 12. We proceed by induction on k and argue by cases. Throughout, $(i \wedge j)$ denotes the MRCA of i and j in $(\mathcal{H}_n)$ and $(i \wedge j)_n$ denotes the MRCA of i and j in $\mathcal{H}_n$.

The first case is that the following event occurs: $\mathcal{T}_k = \mathcal{T}_{k-1}$ and $t_j^{k-1} = t_j^k$ for every $j \geq 1$, or otherwise put, there is $i \in \{-1, \ldots, -k\}$ so that $X_j^i \geq X_j^{k-1}$ for all $j \geq 1$.

Noting that $(i \wedge j)_n \cap (k-1 \wedge j)_n \neq \varnothing$ for sufficiently large n, from (14) and $X_j^i \geq X_j^{k-1}$ we have $(i \wedge j) \subseteq (k-1 \wedge j)$. Thus $i \in (k-1 \wedge j)$, so

$$\{(k-1 \wedge j)_n \cap [n] : j \in [n]\} \subseteq \{(i \wedge j)_n \cap [n] : j \in [n]\},$$

so by Lemma 11,

$$\mathcal{H}_n^{k-1} \subseteq \mathcal{H}_n^i,$$

and it follows that $\mathcal{G}_n^{k-1} = \mathcal{G}_n^k$. Since $\mathcal{I}_n^k = \mathcal{I}_n^{k-1}$ in this case, by the induction hypothesis, $\mathcal{G}_n^k = \mathcal{I}_n^k$.

The second case is that the following event occurs: $\mathcal{T}_k \subsetneq \mathcal{T}_{k-1}$, or otherwise put, for some j, $X_j^{k-1} > \max\{X_j^1, \ldots, X_j^k\}$. It is enough to show two inclusions,

$$\mathcal{I}_n^{k-1} \subseteq \mathcal{G}_n^{k-1} \quad \text{and} \quad \mathcal{H}_n^{k-1} \subseteq \mathcal{I}_n^{k-1}$$

to conclude that $\mathcal{I}_n^{k-1} = \mathcal{G}_n^{k-1}$.

For the first inclusion, we claim that for every point x in the "new branch" $\mathcal{T}_{k-1} \setminus \mathcal{T}_k$, the set $\{j \in [n] : t_j^{k-1} \in F_x(\mathcal{T}_{k-1})\}$ of $\mathcal{I}_n^{k-1}$ is also in $\mathcal{G}_{k-1}$. Therefore fix such x and assume without loss of generality that the set in question is nonempty. If x happens to equal $t_{j_0}^{k-1}$ for some $j_0 \in [n]$ then set $x' = x$, otherwise proceed along the new branch in the 'outward'/away-from-zero/increasing-norm direction until encountering the first sample $t_{j_0}^{k-1}$ with $j_0 \in [n]$, and set $x' = t_j^{k-1}$. More precisely, define x' by

$$x' = \text{ an element of } \{t_j^{k-1} : t_j^{k-1} \in F_x(\mathcal{T}_{k-1}), j \in S_0\} \text{ with minimal } \ell_1\text{-norm},$$

Then $\{j \in [n] : t_j^{k-1} \in F_x(\mathcal{T}_{k-1})\} = \{j \in [n] : t_j^{k-1} \in F_{x'}(\mathcal{T}_{k+1})\}$. According to (23), $\{j \in [n] : t_j^{k-1} \in F_{x'}(\mathcal{T}_{k-1})\} = \{j \in [n] : X_j^{k-1} \geq X_{j_0}^{k-1}\}$ for every element j_0 of $\{t_j^{k-1} : t_j^{k+1} \in F_x(\mathcal{T}_{k-1}), j \in [n]\}$ that has minimal ℓ_1 norm among members of this set. According to Proposition 8(ii), $\{j \in [n] : X_j^{k-1} \geq X_{j_0}^{k-1}\} = (k-1 \cap j_0)_n \cap [n]$, which is an element of $\mathcal{G}_n^{k-1}$. The claim is proved, and in conjunction with the induction hypothesis and Lemma 13 the first inclusion follows.

For the second inclusion, note that if $(k-1 \wedge l)$ contains $i \in \{-1, \ldots, k+1\}$ then $(k-1 \wedge l) \cap [n]$ appears in $\mathcal{G}_n^k$ and hence in $\mathcal{G}_n^{k-1}$ by Lemma 12 and Lemma 13. On the other hand, if $(k-1 \wedge l)$ is disjoint from $\{-1, \ldots, k+1\}$ then

$$\begin{gathered}(k-1 \wedge l) \cap [n] = \{j \in [n] : X_j^k \geq X_j^l\} = \{j \in [n] : t_j^k \in F_x(\mathcal{T}_k)\} \text{ where} \\ x \text{ is the unique point of } \mathcal{T}_k \setminus \mathcal{T}_{k+1} \text{ at distance } X_l^k \text{ from root.}\end{gathered}$$

The second inclusion follows, and we conclude that $\mathcal{I}_n^{k-1} = \mathcal{G}_n^{k-1}$.

The two inclusions taken together show that on the event $\mathcal{T}_k \subsetneq \mathcal{T}_{k+1}$, we have $\mathcal{I}_n^{k-1} = \mathcal{G}_n^{k-1}$ almost surely. This completes the inductive proof.

□

4.3 Part (ii)

Proposition 14. $\mathcal{T}_{-1} \subseteq \mathcal{T}_{-2} \subseteq \ldots$ *almost surely, and the limits*

$$t_j := \lim_{k \to -\infty} t_j^k \quad (j \geq 1)$$

exist almost surely and are members of $\mathcal{T} := cl \bigcup_{k \leq -1} \mathcal{T}_k$*, where cl denotes* ℓ_1*-closure.*

Proof. By (21) the spinal variables (X_j^i) variables take values in $[0,1]$ almost surely. Observe that by definition, $||t_j^k|| = \max\{X_j^{-1}, \ldots, X_j^k\} \leq 1$, and $\pi_{|k|}(t_j^{k-1}) = t_j^k$. The assertions of the proposition follow from definitions and these two facts. $\square$

Let $\mathcal{I}_n$ be the hierarchy derived from $\mathcal{T}$ and the samples $(t_1, \ldots, t_n)$, i.e.

$$\mathcal{I}_n := \{\{j \in [n] : t_j \in F_x(\mathcal{T})\} : x \in \mathcal{T}\} \cup \Xi([n])$$

where $\Xi([n])$ is the trivial hierarchy on $[n]$.

Proposition 15. *For every positive integer n, $\mathcal{I}_n = \mathcal{G}_n$ almost surely.*

We need the following lemma.

Lemma 16. *For every pair u, v of positive integers, $(u \wedge v) \cap \{-1, -2, \ldots\}$ is nonempty with probability one. Here, $(u \wedge v)$ denotes the MRCA of u and v in $(\mathcal{H}_n)$.*

Proof. Define a family $(W_j, j \in \mathbb{Z} \setminus \{i\})$ by

$$W_j = \begin{cases} 1 & \text{if } (u \wedge j) = \{u, j\} \\ 0 & \text{otherwise} \end{cases} \qquad (j \in \mathbb{Z} \setminus \{u, v\}).$$

For distinct integers j_1, j_2, the event $\{W_{j_1} = W_{j_2} = 1\}$ is null set, because $W_{j_1} = W_{j_2} = 1$ implies that $\{j_1, u\}$ and $\{j_2, u\}$ are both members of $\mathcal{H}_n$ for all sufficiently large n, contradicting part (b) of Definition 1.1. Therefore there is almost surely at most one 1 in the sequence (W_j). Since (W_j) is easily seen to be exchangeable, by de Finetti's theorem $W_j = 0$ almost surely for all j.

It follows that $(\mathtt{A}_H(u, v, j), j \in \mathbb{Z} \setminus \{i, j\})$ is a family of Bernoulli variables with at least one 1, almost surely (see (9) for the definition of $\mathtt{A}$). Since $(\mathtt{A}_H(u, v, j))$ is an exchangeable family, the conclusion follows from de Finetti's theorem.

$\square$

Proof of Proposition 15. Set $\mathcal{T}^\circ := \bigcup_{k<-1} \mathcal{T}_k$ and $\partial\mathcal{T} := \mathcal{T} \setminus \mathcal{T}^\circ$. By Proposition 10 and Lemma 13 it follows that

$$\mathcal{G}_n = \bigcup_{k \leq -1} \mathcal{G}_n^k = \bigcup_{k \leq -1} \mathcal{I}_n^k \subseteq \{\{j \in [n] : t_j \in F_x(\mathcal{T})\} : x \in \mathcal{T}^\circ\} \cup \Xi([n])$$

holds for every positive integer n. It remains to establish that

$$\{j \in [n] : t_j \in F_x(\mathcal{T})\} \subseteq \mathcal{G}_n \tag{26}$$

holds for every $x \in \partial\mathcal{T}$. If the set in (26) is empty or a singleton it is in $\mathcal{G}_n$ by definition, therefore without loss of generality suppose $\{u, v\} \subseteq \{j \in [n] : t_j \in F_x(\mathcal{T})\}$ for some distinct pair $u, v \in [n]$ and $x \in \partial\mathcal{T}$. We will derive a contradiction.

We claim first that given these assumptions, $t_u = x = t_v$ almost surely. To see this, note that since $x \in \partial\mathcal{T}$, $x = (x_1, x_2, \ldots)$ does not terminate in zeros, ie. $x_l \neq 0$ infinitely often. It follows that t_u does not terminate in zeros, i.e. $t_u \in \partial\mathcal{T}$, for otherwise x could not be in $[[0, t_u]]_{sp}$. Now, by definitions it follows that the only point of $[[0, t_u]]_{sp}$ that does not terminate in zeros is t_u itself, so since $x \in [[0, t_u]]_{sp}$ (because $t_u \in F_x(\mathcal{T})$) we must have $x = t_u$, and similarly for t_v.

Since $x = t_u = \lim_{k\to-\infty} t_u^k$ is in $\partial\mathcal{T}$, it follows that there is a subsequence k_m of $\{-1,-2,\ldots\}$ for which

$$||t_u^{k_1}|| < ||t_u^{k_2}|| \ldots. \tag{27}$$

Let $(k_m, m \geq 1)$ be the subsequence of $\{-1,-2,\ldots\}$ consisting of the times at which $(X_u^i, i \leq -1)$ exceeds its past maximum,

$$i_1 = -1 \quad i_{m+1} := \max\{i < k_m : X_u^i > X_u^{k_m}\}.$$

Since $||t_u^k|| = \max\{X_u^{-1}, \ldots, X_u^k\}$, this sequence $(k_m, m \leq 1)$ is well-defined. Now by (14) it follows that

$$(k_1 \wedge u) \supsetneq (k_2 \wedge u) \supsetneq \ldots$$

is a strictly decreasing nested family of sets. We claim that $v \in \bigcap_{m\geq1}(k_m \wedge u)$. This is apparent from the proof of Proposition 10, where it is shown that

$$(k_m \wedge u) \cap [n] = \{j \in [n] : t_j^{k_m} \in F_{t_u^{k_m}}(\mathcal{T}_k)\},$$

since $t_v^{k_m} = \pi_{|k_m|}(t_v) = \pi_{|k_m|}(t_u) = t_u^{k_m}$.

Since u and v are both contained in $\bigcap_{m\geq1}(k_m \wedge u)$, it follows that $(u \wedge v) \subseteq \bigcap_{m\geq1}(k_m \wedge u)$. From Proposition 16 there is then with probability one a negative number – we can let i denote the maximum such number – for which $i \in \bigcap_{m\geq1}(k_m \wedge u)$. It follows that $(i \wedge u) \subsetneq (k_m \wedge u)$ for all m, so that $X_u^i \geq X_u^{k_m}$ for every m by (21), contradicting the definition of (k_m).

We have obtained the desired contradiction. It follows that for every fixed $x \in \partial\mathcal{T}$, the set $\{j \in [n] : t_j \in F_x(\mathcal{T})\}$, if nonempty, is with probability one a singleton and therefore an element of $\mathcal{G}_n$. Now observe that

$$\{\{j \in [n] : t_j \in F_x(\mathcal{T})\} : x \in \mathcal{T}\} \cup \Xi([n]) = \{\{j \in [n] : t_j \in F_{t_i}(\mathcal{T})\} : i \in [n]\} \cup \Xi([n]) \quad a.s.$$

Proposition 15 follows. □

Remark. The proof of Proposition 15 shows that the restriction of p to the set $\mathcal{T} \setminus \bigcup_{k<-1} \mathcal{T}_k$ is diffuse, i.e. nonatomic.

4.4 Part (iii)

Proposition 17. *The hierarchies $(\mathcal{G}_n, n \geq 1)$ and $(\mathcal{H}'_n, n \geq 1)$ are equal in distribution.*

It should perhaps be pointed out again that $(\mathcal{H}'_n, n \geq 1)$ is the hierarchy on $\mathbb{N}$ with which we started, i.e. with which we defined the hierarchy $(\mathcal{H}_n)$ on $\mathbb{Z}$. We will need the following lemma:

Lemma 18. *The following equality holds almost surely,*

$$\{(i \wedge j) \cap [\pm n] : j \in \mathbb{Z}, i < 0\} = \{(l \wedge j) \cap [\pm n] : l, j \in \mathbb{Z}\},$$

where $(i \wedge j)$ and $(l \wedge j)$ denote MRCAs in $(\mathcal{H}_n)$.

Proof. We need only prove the "$\supseteq$" direction of the equality. Fix l, j in $\mathbb{Z}$, and let μ_j be the directing measure of the exchangeable sequence $(X_n^j, n \in \mathbb{Z} \setminus \{j\})$. There are three cases to consider.

- X_l^j is an atom of the directing measure μ_j. Then there is almost surely some negative integer i for which $X_i^j = X_l^j$. Then by Proposition 8 part (ii),

$$(i \wedge j) = \{k \in \mathbb{Z} : X_k^j \geq X_i^j\} = \{k \in \mathbb{Z} : X_k^j \geq X_l^j\} = (j \wedge l)$$

  and the claim follows.

- X_l^j is not an atom of μ_j. Recalling the discussion of left-uniformization preceding Theorem 7, it can be seen that with probability 1 there is some negative integer i for which $\max\{X_k^j :\in [\pm n], k \notin (j \wedge l)\} < X_i^j < X_l^j$. Then

$$(i \wedge j) \cap [\pm n] = \{m \in [\pm n] : X_m^j \geq X_i^j\} = \{m \in [\pm n] : X_m^j \geq X_l^j\} = (l \wedge j) \cap [\pm n].$$

The third case, on which we need not linger, is that a probability zero event occurs, e.g. X_l^j lies outside the support of μ_j. □

Lemma 19. *There is the following equality in distribution for all n,*

$$\mathcal{H}'_n \stackrel{d}{=} \{(l \wedge j)_{\mathcal{H}} \cap [n] : l, j \in [n] \cup \Xi([n])$$

where $(l \wedge j)_{\mathcal{H}}$ is the MRCA of l and j in $(\mathcal{H}_n)$.

Proof. Let us say that (16) defines $(\mathcal{H}_n)$ as *image of* $(\mathcal{H}'_n)$ *under* b, and write $(\mathcal{H}_n) = b((\mathcal{H}'_n))$ to express this succinctly. Let c be a bijection from $\mathbb{Z}$ to $\mathbb{Z}$, and let $(\widehat{\mathcal{H}}_n) = c((\mathcal{H}_n))$ be the image of $(\mathcal{H}_n)$ under c. By exchangeability of $(\mathcal{H}'_n)$, there is the following equality in distribution,

$$(\widehat{\mathcal{H}}_n) := c((\mathcal{H}_n)) \stackrel{d}{=} (\mathcal{H}_n) := b((\mathcal{H}'_n))$$

which holds for all fixed bijections $c : \mathbb{Z} \mapsto \mathbb{Z}$ and $b : \mathbb{N} \mapsto \mathbb{Z}$. It follows that

$$\widehat{\mathcal{H}}_n\Big|_{[n]} \stackrel{d}{=} \mathcal{H}_n\Big|_{[n]}.$$

Now choose c so that $c(b(j)) = j$ for $j = 1, \ldots, n$. It is straightforward to check that $\widehat{\mathcal{H}}_n\Big|_{[n]} = \mathcal{H}'_n$ almost surely for this choice of c. Note finally that

$$\mathcal{H}_n\Big|_{[n]} = \{(l \wedge j) \cap [n] : l, j \in [n]\} \cup \Xi([n])$$

by Proposition 5. This establishes the claim. □

Proof of Proposition 17. From Lemma 18, Proposition 5, and Lemma 19 we have

$$\begin{aligned}
\mathcal{G}_n &= \{(i \wedge j)_{\mathcal{H}} \cap [n] : i \leq 0, j \in \mathbb{Z}\} \cup \Xi([n]) \\
&= \{(i \wedge j)_{\mathcal{H}} \cap [n] : i, j \in \mathbb{Z}\} \cup \Xi([n]) \\
&= \{(i \wedge j) \cap [n] : i, j \in [n]\} \cup \Xi([n]) \\
&\stackrel{d}{=} \mathcal{H}'_n,
\end{aligned}$$

where $(i \wedge j)_{\mathcal{H}}$ denotes the MRCA of i and j in $(\mathcal{H}_n)$. □

5 Proof of Theorem 3

Proof. Suppose WLOG that $(\mathcal{H}_n)$ is the hierarchy derived from a real tree $\mathcal{T}$ and an exchangeable family $(t_j, j \geq 1)$ of random elements of $\mathcal{T}$ having directing measure p. We may further suppose that $\mathcal{T}$ is embedded in ℓ_1 by a stick-breaking procedure as in the proof of Theorem 1 or as in Example of Section 3.1. For $k \in \mathbb{N}$ let π_k be the orthogonal projection onto the the span of the first k standard basis elements of ℓ_1,

$$\pi_k((x_1, x_2, \ldots)) = (x_1, \ldots, x_k, 0, 0, \ldots).$$

We will define a map $\xi : [0,1] \mapsto \mathcal{T}$ such that for every $k \geq 1$, and every point $x \in \mathcal{T}_k := \{\pi_k(x) : x \in \mathcal{T}\}$,

(a) $\xi^{-1}(F_x(\mathcal{T}))$ is an interval,

(b) the Lebesgue measure of $\xi^{-1}(F_x(\mathcal{T}))$ equals $p(F_x(\mathcal{T}))$.

To that end, for $k \geq 1$ let p_k be the image of p under π_k. The branches of $\mathcal{T}_k$ can be visualized as strings, and atoms of p_k can be visualized as *beads* on these strings. With this imagery, $(\mathcal{T}_{k+1}, p_{k+1})$ is derived as if by selecting a bead of p_k, crushing this bead into a series of smaller beads, and then drawing these smaller beads out onto the new string $\mathcal{T}_{k+1} \setminus \mathcal{T}_k$, possibly leaving some mass at the location of the crushed atom. (There is also the possibility that $(\mathcal{T}_{k+1}, p_{k+1}) = (\mathcal{T}_k, p_k)$, but this may be ignored.) Let $\xi_1 : [0,1] \mapsto \mathcal{T}_1$ be

- an increasing map, meaning that $x < y$ implies $||\xi_1(x)|| < ||\xi_1(y)||$
- such that p_1 is the image of Lebesgue measure under ξ_1.

Now, $\mathcal{T}_2$ is derived as if by selecting an atom of p_1, crushing it, and stringing the crushed bits in the $\mathbf{e}_2$ direction. If a is the selected atom of p_1, then $\xi_1^{-1}(\{a\})$ is an interval $(u, v]$ or $[u, v]$ in [0,1]. We may therefore define a modification ξ_2 of ξ_1, in such a way that

- ξ_2 agrees with ξ_1 off of $(u, v]$ (or $[u, v]$ as the case may be)
- ξ_2 sends (u, v) onto $a \cup \mathcal{T}_2 \setminus \mathcal{T}_1$
- $u \leq s < t \leq v$ implies $||\xi_2(s)|| \leq ||\xi_2(t)||$
- p_2 is the image of Lebesgue measure under ξ_2.

With this type of construction we can establish the existence of a family of maps $(\xi_k, k \geq 1)$, such that

1. for all $x \in [0,1]$, $\pi_k(\xi_{k+1}(x)) = \xi_k(x)$
2. if $\xi_k(x) = \xi_k(y)$ and $x < y$, then $\xi_{k+1}(x) < \xi_{k+1}(y)$
3. p_k is the image of Lebesgue measure under ξ_k.

It is straightforward to show that the limit $\xi := \lim_{k\to\infty} \xi_k$ exists Lebsegue a.e. and has the asserted properties (a) and (b). Now set

$$\mathscr{H} := \left\{ \xi^{-1}(F_x(\mathcal{T})) : x \in \bigcup_k \mathcal{T}_k \right\} \cup \Xi([0,1]).$$

For (U_j) an IID sequence of uniform[0,1] random variables independent of $\mathcal{T}$ and $n \geq 1$ let

$$\mathcal{H}'_n := \{\{j \in [n] : U_j \in B\} : B \in \mathscr{H}$$

and let

$$\mathcal{H}''_n := \{\{j \in [n] : \xi(U_j) \in F_x(\mathcal{T})\} : x \in \mathcal{T}\} \cup \Xi([n]).$$

It is easily seen that $\mathcal{H}'_n = \mathcal{H}''_n$ almost surely. Also, conditionally given $(\mathcal{T}, p)$, the sequence $(\xi(U_1), \ldots, \xi(U_n))$ is an IID sequence of points with common distribution p. An argument such as can be found in the proof of Proposition 15 shows that if $x \in \mathcal{T} \setminus \bigcup_k \mathcal{T}_k$ then $\{j \in [n] : \xi(U_j) \in F_x(\mathcal{T})\}$ is with probability one either empty or a singleton. It follows that

$$(\mathcal{H}''_n) \stackrel{d}{=} (\mathcal{H}_n).$$

□

6 Complements

6.1 Properties of p and $(\mathcal{H}_n)$

Let $\mathscr{H}$ denote the following class of subsets of the closed interval [0,3],

$$\begin{aligned} \mathscr{H} : &= \{(0,1),(1,2),(2,3)\} \cup \left\{ \bigcup_{n\geq 1} \left\{ \left(\frac{j}{2^n}, \frac{j+1}{2^n} \right) : 0 \leq j \leq 2^n - 1 \right\} \right\} \\ &\cup \{(2,x) : 2 < x < 3\} \cup \Xi([0,3]). \end{aligned}$$

Let $(U_n, n \geq 1)$ be an iid sequence of Uniform[0,3] random variables, and define an exchangeable hierarchy on $\mathbb{N}$ by

$$\mathcal{H}_n := \{\{j \in [n] : U_j \in B\} : B \in \mathscr{H}\}. \tag{28}$$

Figure 4 shows the graph $\mathtt{T}_n$ of $\mathcal{H}_n$ for large n, omitting leaf labels. Let us describe a few key features of this graph $\mathtt{T}_n$ and relate them to $\mathscr{H}$.

- The root of $\mathtt{T}_n$ has degree three. The three vertices v_1, v_2, v_3 connected to the root correspond to the three subintervals $(0,1),(1,2),(2,3)$ of $[0,3]$ contained in $\mathscr{H}$.
- The graph of $\mathtt{T}_n$ exhibits recursive binary splitting below the vertex v_1; this is a consequence of the recursive binary splitting of $(0,1)$ in $\mathscr{H}$.
- The graph of $\mathtt{T}_n$ looks star-like or broomstick-like below v_2; this is because $\mathscr{H}$ contains no nonsingleton subsets of (1,2).

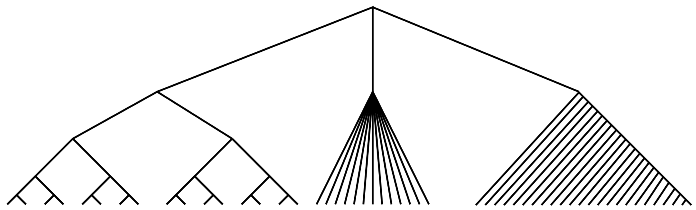

Figure 4: Graph of the hierarchy defined by (28) with leaf labels omitted.

- The graph of $\mathtt{T}_n$ looks like a comb or a caterpillar below v_3; this is because $\mathscr{H}$ contains a family of subsets of $(2, 3)$ of the form $(2, x)$ for x in a dense subset of $(2, 3)$.

From this example, one might make the following naive conjecture.

> **Naive Conjecture:** The three phenomena exhibited by $(\mathcal{H}_n)$ and its graph – infinite recursive splitting, finite splitting, and comblike erosion – are the basic building blocks out of which every exchangeable hierarchy is made.

However, we have difficulty seeing how to make this conjecture more precise: comblike erosion can be interspersed with recursive splitting, splits need not be binary, and a countable family of splits can precede another countable family of splits, and it may be that this latter family of splits is not well-ordered by containment. It is easy to imagine pathological examples of hierarchies.

In lieu of a precise form of the conjecture, we offer the following propositions, which in conjunction with Theorem 1 represent an effort at proving something like the Naive Conjecture. But first, supposing that $(\mathcal{H}_n)$ is a hierarchy on $\mathbb{N}$, we set

$$\alpha_n(j) := \bigcap_{\substack{G \in \mathcal{H}_n \\ j \in G, G \neq \{j\}}} G, \qquad (n \geq j)$$

and call $\alpha_n(j)$ the *parent* of j in $\mathcal{H}_n$; it is easily checked that $\alpha_n(j) \in \mathcal{H}_{n+j}$ for all $n, j \geq 1$.

Also, for $(\mathcal{H}_n)$ a hierarchy on $\mathbb{N}$ and $i, j \in \mathbb{N}$ we write $i \preceq j$ if either $i = j$ or for all $n \geq \max\{i, j\}$,

- $\alpha_n(i) \subsetneq (i \wedge j)_n = \alpha_n(j)$, and
- if u is in $\{i\} \cup (i \wedge j)_n \setminus \alpha_n(i)$ and $v \in [n]$ then

$$\alpha_n(u) = \alpha_n(v) \text{ implies } u = v.$$

Less formally, we write $i \preceq j$ for distinct $i, j \in \mathbb{N}$ if for every $n \geq \max\{i, j\}$, the graph of $\mathcal{H}_n$ looks like a comb in the neighborhood of i and j, and i is "lower down" in this

comb than is j. Next, let Π be the partition of $\mathbb{N}$ derived by putting i and j in the same block if and only if either $i \preceq j$ or $j \preceq i$. We say that Π is the *comb-partition* of $(\mathcal{H}_n)$ and the blocks of Π are the *comb components* of $(\mathcal{H}_n)$. It is easily checked that if $(\mathcal{H}_n)$ is an exchangeable random hierarchy on $\mathbb{N}$ then the comb-partition is exchangeable.

Proposition 20. *Suppose that $(\mathcal{T}, p)$ is a random weighted real tree, that (t_j) an exchangeable sequence directed by p, and that $(\mathcal{H}_n)$ is the exchangeable hierarchy on $\mathbb{N}$ derived from $\mathcal{T}$ and (t_j). On the event that*

- *there is a segment $[[u,v]]$ of $\mathcal{T}$ that is oriented towards the root of $\mathcal{T}$, with v further from the root, meaning that $[[u,v]] \subseteq [[0,v]]$*
- *such that $[[u,v]]$ does not sprout any branches of positive p-mass, meaning that for all x in the support of p with $x \notin [[u,v]]$,*

$$u \in [[0,x]] \text{ implies } v \in [[0,x]]$$

- *such that $p_a([[u,v]]) = 0$ and $p_d([[u,v]]^\circ) > 0$, where p_a and p_d are the atomic and diffuse components of p, respectively, and $[[u,v]]^\circ$ denotes the interior of $[[u,v]]$ for the topology of $\mathcal{T}$*

the set $\{j : t_j \in [[u,v]]^\circ\}$ is a subset of one of the comb-components of $(\mathcal{H}_n)$. Conversely, on the event that distinct positive integers i and j lie in the same comb-component of $(\mathcal{H}_n)$, there is with probability one a segment $[[u,v]]$ of $\mathcal{T}$ having the properties above for which $t_i, t_j \in [[u,v]]^\circ$.

The statement of the proposition is obvious from definitions.

Proposition 21. *Suppose that $(\mathcal{T}, p)$ is a random weighted real tree, that (t_j) an exchangeable sequence directed by p, and that $(\mathcal{H}_n)$ is the exchangeable hierarchy on $\mathbb{N}$ derived from $\mathcal{T}$ and (t_j). On the event that a is an atom of p, for all distinct pairs u, v in the set $B = \{j : t_j = a\}$, $(u \wedge v) = B$, where $(u \wedge v)$ denotes the MRCA of u and v in the hierarchy derived from (t_j) and $\mathcal{T}$. Furthermore, t_j is an atom of p if and only if*

$$0 < \lim_{n\to\infty} \frac{1}{n} \#\{k \in [n] : \alpha_m(j) = \alpha_m(k) \text{ for all } m \geq \max\{j,k\}\},$$

and if t_j is an atom of p then $p(\{t_j\})$ equals this limit above almost surely.

The proof of this proposition is elementary and is therefore omitted.

Remark. The atomic and diffuse parts p_a and p_d of the random measure p of Theorem 1 are "invariants," loosely speaking, of the exchangeable hierarchy $(\mathcal{H}_n)$ of that theorem. More formally, p_a and p_d are measurable functions of p, within the standard abstract setup for random measures [53, Chapter 1].

6.2 EPPFs, EHPFs, etc.

Throughout this section, let $\mathtt{C} := \bigcup_{k\geq 1} \mathbb{N}^k$ denote the set of compositions of positive integers.

Suppose that Π is an exchangeable random partition of $\mathbb{N}$ and for $n \geq 1$ let $\Pi_n = \Pi\Big|_{[n]}$ denote the the restriction of Π to $[n]$. Then it is straightforward to show that there is a symmetric function $p : \mathtt{C} \mapsto [0,1]$ such that for any partition $\pi_n = \{B_1, \ldots, B_k\}$ of $[n]$ into disjoint blocks $B_1, \ldots, B_k$,

$$\mathbb{P}(\Pi_n = \pi_n) = p(\#B_1, \ldots, \#B_k) \tag{29}$$

where $(\#B_1, \ldots, \#B_k)$ are the sizes of the blocks of π_n. Both the number of blocks and the sequence of block sizes can be regarded as functions of π_n. Blocks of a partition are conventionally ordered by least elements, or alternatively by size, but this is immaterial for the present discussion because symmetry of p means that the order in which block sizes are presented does not matter. More formally, symmetry of p means that for every $k \geq 1$,

$$p(\lambda_1, \ldots, \lambda_k) = p(\lambda_{\sigma(1)}, \ldots, \lambda_{\sigma(k)}) \tag{30}$$

for every element $(\lambda_1, \ldots, \lambda_k)$ of $\mathtt{C}$ and every permutation σ of $[k]$. Another property of p comes from the consistency of Π_n as n varies: by exchangeability, the distribution of Π_n is the same as the distribution of the partition derived from Π_{n+1} by first relabeling the contents of the blocks of Π_{n+1} using a uniform random permutation of $[n+1]$ and then restricting the resulting partition to $[n]$. In purely algebraic terms this translates as the following *addition rule*:

$$p(\lambda_1, \ldots, \lambda_k) = p(\lambda_1 + 1, \ldots, \lambda_k) + \ldots + p(\lambda_1, \ldots, \lambda_k + 1) + p(\lambda_1, \ldots, \lambda_k, 1). \tag{31}$$

There is also the trivial normalization condition

$$p(1) = 1. \tag{32}$$

We say that $p : \mathtt{C} \mapsto [0,1]$ is an *exchangeable partition probability function* (EPPF) if p satisfies (30) - (32); an easy construction shows that if p is an EPPF then there is an exchangeable random partition Π of $\mathbb{N}$ for which (29) holds for all $n \geq 1$. EPPFs are therefore in 1-1 correspondence with distributions of exchangeable partitions. For more on EPPFs see [63] and [62, Chapters 2 & 3].

According to a result in [63], if p is an EPPF then there is a sequence $(P_1, P_2, \ldots)$ of nonnegative random variables for which

$$p(\lambda_1, \ldots, \lambda_k) = \mathbb{E}\left[\left(\prod_{i=1}^{k} P_i^{\lambda_i - 1}\right)\left(\prod_{i=2}^{k}(1 - P_1 - \ldots - P_{i-1})\right)\right] \tag{33}$$

holds for all $\lambda \in \mathtt{C}$. The EPPF p determines the joint law of this sequence (P_i) uniquely, and conversely. There is also the following consequence of Theorem 2: if Π is an exchangeable random partition of $\mathbb{N}$ for which the sequence (P_i) of (33) satisfies

$$\mathbb{P}\left(M := \inf\{n : P_n = 0\} > 1 \text{ and if } M < \infty \text{ then } P_{M+i} = 0 \text{ for all } i \geq 0\right) = 1$$

or equivalently, if Π is an exchangeable random partition of $\mathbb{N}$ that almost surely does not contain any singleton blocks, then there is a measure μ on the Kingman simplex $\nabla := \{(x_1, x_2, \ldots) : x_1 \geq x_2 \geq \ldots, \sum_{i \geq 1} x_i \leq 1\}$ for which for all $\lambda \in \bigcup_{k \geq 1} \mathbb{N}^k$,

$$p(\lambda) = \int_{\nabla} m_\lambda(x) \mu(dx), \tag{34}$$

where m_λ is the monomial symmetric polynomial

$$m_\lambda(x) := \sum_\sigma x_{\sigma(1)}^{\lambda_1} \dots x_{\sigma(k)}^{\lambda_k} \tag{35}$$

for $\lambda = (\lambda_1, \dots, \lambda_k)$, where the sum is taken over all injective functions $\sigma : [k] \mapsto \mathbb{N}$. The relation between the (P_i) of (33) and the measure μ of (34) can be succinctly described: μ is the distribution of the rearrangement of (P_i) in nonincreasing order. It can therefore be seen that this μ is therefore supported on the set

$$\{x \in \nabla : \sum_i x_{i \geq 1} = 1\}. \tag{36}$$

For proofs of the material in this exposition, see [63] and [62, Chapters 2 & 3].

It is natural to ask whether there is an analogous story for exchangeable hierarchies on $\mathbb{N}$. We are unable to find a "hierarchies counterpart" to (33), i.e. a *moment formula* which relates *every* exchangeable hierarchy to a the moments of a family of random variables. But the remaining formulae above all have analogues in the exchangeable hierarchies setting. Making these analogies explicit involves the use of a new (to us) family of symmetric polynomials.

Recall from the introduction that the *graph* of a hierarchy $\mathcal{H}_n$ on $[n]$ is a rooted tree $\mathtt{T}_n$ with n leaves, where each leaf bears a distinct label in $[n]$. (The graph of a hierarchy also lacks non-root internal vertices of degree two, and furthermore lacks edge lengths and orientation of edges at vertices vertex.) We define the *shape* of such a tree $\mathtt{T}_n$ to be its orbit under the action of the symmetric group, the action of a permutation σ being to relabel leaf i by $\sigma(i)$ for every $i \in [n]$. We use lower-case bold face $\mathtt{s}_n$ to denote the shape of $\mathtt{T}_n$, which can be regarded as a function of $\mathtt{T}_n$ or alternatively of $\mathcal{H}_n$, $\mathtt{s}_n = \mathtt{s}(\mathtt{T}_n) = \mathtt{s}(\mathcal{H}_n)$. Obviously, the shape of $\mathtt{T}_n$ can be identified with the unlabeled tree derived by erasing the labels on the leaves of $\mathtt{T}_n$, but this observation is not far from a tautology, as unlabeled graphs are often defined as orbits under such actions of the symmetric group.

We write $\mathtt{s}_n \nearrow \mathtt{s}_{n+1}$ if there is a hierarchy $\mathcal{H}_{n+1}$ on $[n+1]$ for which $\mathtt{s}_n = \mathtt{s}(\mathcal{H}_{n+1}\big|_{[n]})$ and $\mathtt{s}_{n+1} = \mathtt{s}(\mathcal{H}_{n+1})$. Equivalently (regarding the shapes $\mathtt{s}_n$ and $\mathtt{s}_{n+1}$ as unlabeled graphs), we write $\mathtt{s}_n \nearrow \mathtt{s}_{n+1}$ if it is possible to remove a leaf of $\mathtt{s}_{n+1}$ and thereby obtain $\mathtt{s}_n$. In this context, *removing a leaf* means (i) erasing the leaf and the edge of $\mathtt{s}_{n+1}$ which had that leaf as an endpoint and then (ii) erasing any nonroot internal vertices of the resulting graph that happen to have degree two – such a vertex appears if the parent of the erased leaf happens to have degree 3 in $\mathtt{s}_{n+1}$.

Let $\mathtt{S} := \{\mathtt{s}(\mathcal{H}_n) : \mathcal{H}_n \text{ a hierarchy on } [n] \text{ for some } n \geq 1\}$. With these definitions the following analogues of (30) - (32) are obvious: if $(\mathcal{H}_n)$ is an exchangeable hierarchy on $\mathbb{N}$, then there is a function $h : \mathtt{S} \mapsto [0,1]$ for which for every fixed hierarchy H on $[n]$,

$$\mathbb{P}(\mathcal{H}_n = H) = h(\mathtt{s}(H)) \tag{37}$$

$$h(\mathtt{s}_n) = \sum_{\mathtt{s}_{n+1} : \mathtt{s}_n \nearrow \mathtt{s}_{n+1}} h(\mathtt{s}_{n+1}) \tag{38}$$

$$h\Big(\mathtt{s}(\Xi([1]))\Big) = 1. \tag{39}$$

Here, $\Xi([1])$ is the trivial hierarchy on $[1]$. Note that (29) and (30) together say that $\mathbb{P}(\Pi_n = \pi_n)$ is a function of the equivalence class of π_n, two partitions of $[n]$ being equivalent if and only if they have the same block sizes. Formula (37) likewise asserts that $\mathbb{P}(\mathcal{H}_n = H)$ only depends on the equivalence class of H, i.e. the shape of the graph of H. To strengthen the analogy between (38) and (31) we may write $\lambda \nearrow \Lambda$ for $\lambda, \Lambda \in \mathtt{C}$ to mean

$$\lambda = (\lambda_1, \dots, \lambda_k) \text{ and } \Lambda \in \{(\lambda_1 + 1, \dots, \lambda_k), (\lambda_1, \dots, \lambda_k + 1), (\lambda_1, \dots, \lambda_k, 1)\}$$

and rewrite (31) accordingly. We remark that the two relations on $\mathtt{S}$ and on $\mathtt{C}$ that are both denoted $\nearrow$, both turn these respective spaces into graded graphs or lattices in the sense of [55], and that the problem characterizing exchangeable hierarchies on $\mathbb{N}$ (solved by Theorem 1 of the present work) and the problem of characterizing exchangeable partitions of $\mathbb{N}$ (solved by Kingman's Theorem 2), are equivalent to characterizing the classes of bounded, positive harmonic functions on these lattices. See [55, Chapter 0] for much more on this topic.

Proposition 22. *Suppose that $h : \mathtt{S} \mapsto [0,1]$ satisfies (38) for all $n \geq 1$ and (39). Then there is an exchangeable random hierarchy $(\mathcal{H}_n)$ on $\mathbb{N}$ for which (37) holds for every fixed hierarchy H on $[n]$, for all $n \geq 1$.*

Proof. Let $\mathcal{H}_1 = \Xi([1])$, and assuming that $\mathcal{H}_1, \dots, \mathcal{H}_n$ have been defined, conditionally given $\mathcal{H}_n = H$, select $\mathcal{H}_{n+1}$ from the set

$$\left\{ \text{hierachies } H' \text{ on } [n+1] : H'\Big|_{[n]} = H \right\},$$

selecting H' with probability $h(H')/h(H)$. $\square$

As previously remarked, we have no analogue of (33), but we can write down a hierarchies counterpart to (34). Towards this end, we introduce the following family of symmetric polynomials. First, let $\mathbb{R}[x_\lambda, \lambda \in \mathtt{C}]$ denote the ring of formal power series in (commuting) variables x_λ for $\lambda \in \mathtt{C}\}$, with coefficients in $\mathbb{R}$. Then for every function $\sigma : \mathbb{N} \mapsto \mathbb{N}$ let x_σ denote the following generalized monomial,

$$x_\sigma := \prod_{n=1}^{\infty} x_{(\sigma(1), \dots, \sigma(n))}. \tag{40}$$

Next, if Σ is a mapping from $[k]$ to $\mathbb{N}^{\mathbb{N}}$, which is to say, if $\Sigma_i : \mathbb{N} \mapsto \mathbb{N}$ for $i \in [k]$, then let the *class* of Σ be the following hierarchy on $[k]$,

$$\text{class}(\Sigma) := \{\{j \in [k] : (\Sigma_j(1), \dots, \Sigma_j(n)) = \lambda\} : n \geq 1 \text{ and } \lambda \in \mathbb{N}^n\} \cup \Xi([k]).$$

Finally, for H a fixed hierarchy on $[k]$, let m_H be the following symmetric polynomial,

$$m_H(x) := \sum_{\Sigma : \text{class}(\Sigma) = H} x_{\Sigma_1} \dots x_{\Sigma_k},$$

where the sum is taken over all $\Sigma \in [k]^{\mathbb{N}^{\mathbb{N}}}$. To see how these notions might be useful, suppose that $\mathscr{H}$ is a nonrandom hierarchy on [0,1] consisting of nested intervals: at the

top level of $\mathscr{H}$, [0,1] is partitioned into intervals of lengths $w_1, w_2, \ldots$, which sum up to 1, where we list these lengths in nonincreasing order of size. Then for every $i \geq 1$ the interval w_i is partitioned into further subintervals of lengths $w_{(i,1)}, w_{(i,2)}, \ldots$, which sum up to w_i, and these lengths are likewise ordered by size. We suppose that this recursive process of splitting continues indefinitely: the intervals at depth D have widths w_λ for $\lambda \in \mathbb{N}^D$, and every interval of width w_λ is split into further subintervals of widths $w_{\lambda,1}, w_{\lambda,2}, \ldots$, which sum up to w_λ, and are ordered by size. We then set $x_{(n)} := w_{(n)}$ for all $n \geq 1$ and

$$x_\lambda := \frac{w_\lambda}{w_{\lambda'}} \qquad (\text{length}(\lambda) \geq 2)$$

where λ' is the composition derived by removing the final element of λ. The family $\{x_\lambda, \lambda \in \mathtt{C}\}$ then has the following two properties,

(i) $\sum_{n\geq 1} x_{(n)} = 1$ and for every $\lambda \in \mathtt{C}$, $\sum_{n\geq 1} x_{(\lambda,n)} = 1$

(ii) $x_{(1)} \geq x_{(2)} \geq \ldots$ and $x_{(\lambda,1)} \geq x_{(\lambda,2)} \geq \ldots$.

Note that for every $\lambda \in \mathtt{C}$,

$$w_\lambda = \sum_{\substack{\sigma:\mathbb{N}\mapsto\mathbb{N} \\ (\sigma(1),\ldots,\sigma(n))=\lambda}} x_\sigma$$

for x_σ the monomial defined by (40). Now let $(U_1, \ldots, U_n)$ be iid uniform[0,1] random variables, and define a hierarchy $\mathcal{H}_n$ on $[n]$ by

$$\mathcal{H}_n := \{\{j \in [n] : U_j \in B\} : B \in \mathscr{H}\}.$$

Then for every fixed hierarchy H on $[n]$,

$$\mathbb{P}(\mathcal{H}_n = H) = m_H(x).$$

If if the interval hierarchy $\mathscr{H}$ is random then there is a probability measure μ on $\{x_\lambda, \lambda \in \mathtt{C}\}$ for which items (i) and (ii) above hold μ-a.s., and in this case

$$\mathbb{P}(\mathcal{H}_n = H) = h(\mathtt{s}(H)) = \int_{\nabla*} m_H(x)\mu(dx), \tag{41}$$

where $\nabla*$ denotes the subset of $\mathtt{C}^\mathbb{R}$ for which for every $x \in \mathtt{C}^\mathbb{R}$,

(i) $\sum_{n\geq 1} x_{(n)} \leq 1$ and for every $\lambda \in \mathtt{C}$, $\sum_{n\geq 1} x_{(\lambda,n)} \leq 1$

(ii) $x_{(1)} \geq x_{(2)} \geq \ldots$ and $x_{(\lambda,1)} \geq x_{(\lambda,2)} \geq \ldots$.

Equation (41) is our analogue of (34). We make no claims on rigor, here; topological properties, convergence of distributions, and analytic properties of symmetric functions have been well-studied for the Kingman simplex ∇ [27, 51, 39], but we have not even defined a measurable structure on $\nabla*$. We leave such issues for another paper.

Equation (34) does not describe the distribution of the most general exchangeable partition, and likewise (41) does not describe the distribution of the most general exchangeable hierarchy. We give two examples of exchangeable hierarchies $(\mathcal{H}_n)$ on $\mathbb{N}$ for which (41) does not describe the distribution of $(\mathcal{H}_n)$. Throughout, (U_n) is an IID family of uniform[0,1] random variables.

(a) For $n \geq 1$ set $\mathcal{H}_n := \{\{j \in [n] : U_j \geq x\} : 0 \leq x \leq 1\}$.

(b) Let (q_n) be an enumeration of the rational numbers in [0,1], and let (ϵ_n) be a sequence of positive numbers summing to 1/4. Let $\mathscr{U}$ be the open subset of [0,1] defined by $\mathscr{U} := \bigcup_{n\geq 1}(q_n - \epsilon_n, q_n + \epsilon_n)$. Then set

$$\mathcal{H}_n := \{\{j \in [n] : U_j \geq x\} : x \in \mathcal{U}^c\} \cup \Xi([n])$$

Speaking loosely, (a) is problematic because there is "continuous erosion", for which (41) cannot account, and (b) is problematic because (41) assumes that the "recursive splits" are well-ordered by inclusion. However, the requirement that the interval hierarchy $\mathscr{H}$ exhibit infinite recursive splitting can be overcome: we can model hierarchies $\mathscr{H}$ that stop splitting at finite depth by setting $w_\lambda = w_{(\lambda,1)} = w_{(\lambda,1,1)}$ and $w_{(\lambda,n)} = 0$ for all $n \geq 2$ for some λ, for example.

We end this section by observing the following multiplication rule for our symmetric polynomials. Recall that for the usual monomial symmetric functions, there is the multiplication rule

$$m_1(x)m_\lambda(x) = \sum_{\Lambda:\lambda\nearrow\Lambda} m_\Lambda(x)$$

which holds for all $\lambda \in \mathtt{C}$. Since $m_1(x) = 1$ on the set (36), this multiplication rule implies that for p *defined* by (34), p satisfies (31). The analogue in our context is the following: for any hierarchy H on $[n]$, and with m_1 denoting the polynomial

$$m_1(x) := m_{\{\{1\},\varnothing\}} = \sum_{\sigma\in\mathbb{N}^{\mathbb{N}}} x_\sigma$$

there is the following multiplication rule,

$$m_1 m_H(x) = \sum_{H_{n+1}:H_{n+1}\big|_{[n]}=H_n} m_{H_{n+1}}(x).$$

Since $m_1(x) = 1$ on the set $\{x \in \nabla* : x \text{ satisfies (i) and (ii) directly above}\}$, this multiplication rule likewise implies that for h defined by the second equality of (41), h satisfies (38).

6.3 Tail measurability and open problems

If $(\mathcal{H}_n)$ is a random hierarchy on $\mathbb{N}$, define the *tail sigma field* of $(\mathcal{H}_n)$ as follows,

$$\text{tail}(\mathcal{H}_n) = \bigcap_{n\geq 1} \sigma\left(\mathcal{H}_{n+1}\Big|_{\{n+1\}}, \mathcal{H}_{n+2}\Big|_{\{n+1,n+2\}}, \ldots\right).$$

If $(\mathcal{H}_n)$ is an exchangeable hierarchy, then the pair $(\mathcal{T}, p)$ of Theorem 1 is not tail-measurable for the following reason. Let $\pi_1(x) = (x_1, 0, 0, \ldots)$ for $x \in \ell_1$. Then the image of p under π_1 is the directing measure for the exchangeable spinal variables $(X_{b^{-1}(j)}^{b^{-1}(1)})$ defined by (21), where b is the bijection mentioned at the beginning of Section 4, but neither these spinal variables nor their directing measure are tail measurable. On the other hand,

Proposition 23. *The distribution of the pair $(\mathcal{T}, p)$ of Theorem 1 is measurable with respect to the exchangeable sequence $(\mathcal{H}_n)$ of that theorem.*

Proof. This is a direct consequence of the fact that the bijection b mentioned at the beginning of Section 4 can be chosen arbitrarily. $\square$

Instead of proving the assertion that $(\mathcal{T}, p)$ is not tail$(\mathcal{H}_n)$-measurable, we offer the following analogy using exchangeable partitions. Suppose that $\mathscr{U}$ is a random open subset of $[0,1]$ having Lebesgue measure one, and let (U_n) and (V_n) be independent IID sequences of uniform[0,1] random variables, jointly independent of $\mathscr{U}$. Form an exchangeable partition Π of $\mathbb{N}$ by putting i and j in the same block of Π if U_i and U_j fall in the same connected component of $\mathscr{U}$, and index the blocks $\{B_1, B_2, \ldots\}$ of Π by least elements, so that $1 = \min B_1 < \min B_2 < \ldots$. Then the limits

$$P_i = P_i(\Pi) = \lim_{n\to\infty} \frac{1}{n} \# B_i \cap [n]$$

exist almost surely, and P_i is the width of the interval of $\mathscr{U}$ containing $U_{\min B_i}$. Now form another open subset $\mathscr{U}'$ by placing intervals of widths P_i in left-to-right order,

$$\mathscr{U}' := (0, P_1) \cup \bigcup_{n\geq 1} (P_1 + \ldots, P_n, P_1 + \ldots + P_{n+1}).$$

Then $\mathscr{U}'$ is not measurable with respect to the tail of Π,

$$\text{tail}(\Pi) = \bigcap_{n\geq 1} \sigma\left(\Pi\Big|_{\{n,n+1,\ldots\}}\right),$$

because $P_1(\Pi)$ is not measurable with respect to tail(Π), but $P_1(\Pi)$ equals almost surely the length of the connected component of $\mathscr{U}'$ whose left-endpoint is zero. Our weighted tree $(\mathcal{T}, p)$ is very much like the open subset $\mathcal{U}'$.

Continuing this discussion, it is evident that if $(P_1^\downarrow, P_2^\downarrow, \ldots)$ is the sequence of P_i's ranked in nonincreasing order, and

$$\mathscr{U}_{\text{ranked}} := (0, P_1^\downarrow) \cup \bigcup_{n\geq 1} (P_1^\downarrow + \ldots, P_n^\downarrow, P_1^\downarrow + \ldots + P_{n+1}^\downarrow), \tag{42}$$

then $\mathscr{U}_{\text{ranked}}$ is measurable with respect to tail(Π). Deterministically reranking the components of $\mathscr{U}'$ in nonincreasing order effectively erases the information contained in $\mathscr{U}'$ but not contained in tail(Π). Obviously, the fact that the resulting order is by decreasing length is immaterial; any deterministic ordering will do.

We may now state the previously-promised stronger version of Kingman's theorem and some open problems.

Theorem 24 ([56])**.** *Suppose that Π is an exchangeable partition of $\mathbb{N}$ and that the probability space supports a sequence (U_n) of IID uniform[0,1] random variables independent of Π. Then there is a Π-measurable random open subset $\mathscr{U}$ of [0,1] such that if Π' is the partition defined by*

$$\{i \text{ and } j \text{ in same block of } \Pi'\} = \{U_i \text{ and } U_j \text{ in same component of } \mathscr{U}, \text{ or } i = j\}.$$

then there is the equality of joint distributions

$$(\Pi, \mathscr{U}) \stackrel{d}{=} (\Pi', \mathscr{U}) \tag{43}$$

Question 1 Define an equivalence relation $\sim$ on laws of weighted real trees, writing $\mathscr{L}(\mathcal{T}, p) \sim \mathscr{L}(\mathcal{T}', p')$ if and only if $(\mathcal{H}_n) \stackrel{d}{=} (\mathcal{H}'_n)$ for $(\mathcal{H}_n)$ and $(\mathcal{H}'_n)$ exchangeable hierarchies derived by sampling from $(\mathcal{T}, p)$ and $(\mathcal{T}', p')$, respectively.

Is there a nice way of telling whether or not $\mathscr{L}(\mathcal{T}, p) \sim \mathscr{L}(\mathcal{T}', p')$? Speaking loosely, it should be possible to *prune away* tree branches of $\mathcal{T}$ that carry no p-mass, and also stretch segments of $\mathcal{T}$ arbitrarily, and not change the equivalence class of $\mathscr{L}(\mathcal{T}, p)$. Purely topological considerations are not quite enough to settle this question: suppose that $\mathcal{T}_1$ is the tree [0,1] rooted at 1 and p_1 is Lebesgue measure on [0,1], and suppose that $\mathcal{T}_2$ is is the half line $[0, \infty)$ rooted at 0, and p_2 is the exponential(1) distribution on $\mathcal{T}_2$. Then $\mathcal{T}_1$ and $\mathcal{T}_2$ are not homeomorphic, but $\mathscr{L}(\mathcal{T}, p) \sim \mathscr{L}(\mathcal{T}', p')$.

Question 2 Is there a nice way to select from each equivalence class of $\sim$ above a unique representative of that equivalence class? Such a recipe would be akin to reordering component intervals of open subsets of [0,1], as discussed above. By nice we mean *measurable*, and you can pick the sigma fields, but the goal is to have an analogy of the strong version of Kingman's theorem involving an equality of joint distributions, as in (43).

Question 3 Repeat the previous questions in the context of Theorem 3, i.e. with hierarchies on [0,1] instead of weighted real trees in ℓ_1.

We thank David Aldous, Steve Evans, and Matthias Winkel for helpful discussion.

References

[1] R. Adams, Z. Ghahramani, and M. I. Jordan, *Tree-structured stick breaking for hierarchical data*, In J. Shawe-Taylor, R. Zemel, J. Lafferty, and C. Williams (Eds.) Advances in Neural Information Processing Systems (NIPS) 23, to appear.

[2] David Aldous, *The continuum random tree. III*, Ann. Probab. **21** (1993), no. 1, 248–289. MR1207226

[3] ———, *Brownian excursions, critical random graphs and the multiplicative coalescent*, Ann. Probab. **25** (1997), no. 2, 812–854. MR1434128

[4] David Aldous, Maxim Krikun, and Lea Popovic, *Stochastic models for phylogenetic trees on higher-order taxa*, J. Math. Biol. **56** (2008), no. 4, 525–557. MR2367031

[5] David Aldous and Jim Pitman, *A family of random trees with random edge lengths*, Random Structures Algorithms **15** (1999), no. 2, 176–195. MR1704343

[6] David J. Aldous, *Representations for partially exchangeable arrays of random variables*, J. Multivariate Anal. **11** (1981), no. 4, 581–598. MR637937

[7] ———, *Exchangeability and related topics*, École d'été de probabilités de Saint-Flour, XIII—1983, Lecture Notes in Math., vol. 1117, Springer, Berlin, 1985, pp. 1–198. MR883646

[8] ———, *Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today*, Statist. Sci. **16** (2001), no. 1, 23–34. MR1838600

[9] Tim Austin, *On exchangeable random variables and the statistics of large graphs and hypergraphs*, Probab. Surv. **5** (2008), 80–145. MR2426176

[10] Julien Berestycki, *Ranked fragmentations*, ESAIM Probab. Statist. **6** (2002), 157–175 (electronic). MR1943145 (2004d:60196)

[11] Julien Berestycki, Nathanaël Berestycki, and Jason Schweinsberg, *Small-time behavior of beta coalescents*, Ann. Inst. Henri Poincaré Probab. Stat. **44** (2008), no. 2, 214–238. MR2446321

[12] Jean Bertoin, *A fragmentation process connected to Brownian motion*, Probab. Theory Related Fields **117** (2000), no. 2, 289–301. MR1771665

[13] ———, *Eternal additive coalescents and certain bridges with exchangeable increments*, Ann. Probab. **29** (2001), no. 1, 344–360. MR1825153

[14] ———, *Homogeneous fragmentation processes*, Probab. Theory Related Fields **121** (2001), no. 3, 301–318. MR1867425

[15] ———, *Self-similar fragmentations*, Ann. Inst. H. Poincaré Probab. Statist. **38** (2002), no. 3, 319–340. MR1899456

[16] ———, *Random covering of an interval and a variation of Kingman's coalescent*, Random Structures Algorithms **25** (2004), no. 3, 277–292. MR2086161

[17] ———, *Random fragmentation and coagulation processes*, Cambridge Studies in Advanced Mathematics, vol. 102, Cambridge University Press, Cambridge, 2006. MR2253162

[18] ———, *Homogeneous multitype fragmentations*, In and out of equilibrium. 2, Progr. Probab., vol. 60, Birkhäuser, Basel, 2008, pp. 161–183. MR2477381

[19] Jean Bertoin and Jean-François Le Gall, *Stochastic flows associated to coalescent processes*, Probab. Theory Related Fields **126** (2003), no. 2, 261–288. MR1990057

[20] Jean Bertoin and Alain Rouault, *Discretization methods for homogeneous fragmentations*, J. London Math. Soc. (2) **72** (2005), no. 1, 91–109. MR2145730

[21] D. M. Blei, T. L. Griffiths, and M. I. Jordan, *The nested chinese restaurant process and bayesian nonparametric inference of topic hierarchies*, Journal of the ACM **57** (2010), 1–30.

[22] C. Blundell, Y. W. Teh, and K. A. Heller, *Bayesian rose trees*, Proceedings of the International Conference on Uncertainty in Artificial Intelligence, 2010.

[23] P. Diaconis and D. Freedman, *de Finetti's theorem for Markov chains*, Ann. Probab. **8** (1980), no. 1, 115–130. MR556418

[24] ———, *Finite exchangeable sequences*, Ann. Probab. **8** (1980), no. 4, 745–764. MR577313

[25] Persi Diaconis and Svante Janson, *Graph limits and exchangeable random graphs*, Rend. Mat. Appl. (7) **28** (2008), no. 1, 33–61. MR2463439

[26] Peter Donnelly, Steven N. Evans, Klaus Fleischmann, Thomas G. Kurtz, and Xiaowen Zhou, *Continuum-sites stepping-stone models, coalescing exchangeable partitions and random trees*, Ann. Probab. **28** (2000), no. 3, 1063–1110. MR1797304

[27] Peter Donnelly and Paul Joyce, *Continuity and weak convergence of ranked and size-biased permutations on the infinite simplex*, Stochastic Process. Appl. **31** (1989), no. 1, 89–103. 996613 (90e:60012)

[28] ———, *Consistent ordered sampling distributions: characterization and convergence*, Adv. in Appl. Probab. **23** (1991), no. 2, 229–258. MR1104078 (92k:60035)

[29] Peter Donnelly and Thomas G. Kurtz, *Particle representations for measure-valued population models*, Ann. Probab. **27** (1999), no. 1, 166–205. MR1681126

[30] Thomas Duquesne and Jean-François Le Gall, *Random trees, Lévy processes and spatial branching processes*, Astérisque (2002), no. 281, vi+147. MR1954248

[31] Richard Durrett, *Probability: theory and examples*, third ed., Thomson, Belmont, CA, 2005. durrett

[32] Steven N. Evans, *Kingman's coalescent as a random metric space*, Stochastic models (Ottawa, ON, 1998), CMS Conf. Proc., vol. 26, Amer. Math. Soc., Providence, RI, 2000, pp. 105–114. MR1765005

[33] ———, *Probability and real trees*, Lecture Notes in Mathematics, vol. 1920, Springer, Berlin, 2008, Lectures from the 35th Summer School on Probability Theory held in Saint-Flour, July 6–23, 2005. MR2351587

[34] David Freedman and Persi Diaconis, *de Finetti's theorem for symmetric location families*, Ann. Statist. **10** (1982), no. 1, 184–189. MR642729

[35] Alexander Gnedin, Alex Iksanov, and Martin Möhle, *On asymptotics of exchangeable coalescents with multiple collisions*, J. Appl. Probab. **45** (2008), no. 4, 1186–1195. MR2484170

[36] Alexander Gnedin and Grigori Olshanski, *A q-analogue of de Finetti's theorem*, Electron. J. Combin. **16** (2009), no. 1, Research Paper 78, 16. MR2529787

[37] Alexander V. Gnedin, *On a class of exchangeable sequences*, Statist. Probab. Lett. **25** (1995), no. 4, 351–355. MR1363235

[38] ———, *The representation of composition structures*, Ann. Probab. **25** (1997), no. 3, 1437–1450. MR1457625

[39] ______, *On convergence and extensions of size-biased permutations*, J. Appl. Probab. **35** (1998), no. 3, 642–650. 1659532 (2000a:60078)

[40] Christina Goldschmidt and James B. Martin, *Random recursive trees and the Bolthausen-Sznitman coalescent*, Electron. J. Probab. **10** (2005), no. 21, 718–745 (electronic). MR2164028

[41] Bénédicte Haas and Grégory Miermont, *The genealogy of self-similar fragmentations with negative index as a continuum random tree*, Electron. J. Probab. **9** (2004), no. 4, 57–97 (electronic). MR2041829

[42] Bénédicte Haas, Grégory Miermont, Jim Pitman, and Matthias Winkel, *Continuum tree asymptotics of discrete fragmentations and applications to phylogenetic models*, Ann. Probab. **36** (2008), no. 5, 1790–1837. MR2440924 (2010d:60192)

[43] Bénédicte Haas, Jim Pitman, and Matthias Winkel, *Spinal partitions and invariance under re-rooting of continuum random trees*, Ann. Probab. **37** (2009), no. 4, 1381–1411. MR2546748

[44] K. A. Heller and Z. Ghahramani, *Bayesian hierarchical clustering*, In Proceedings of the 22nd International Conference on Machine Learning, 2005.

[45] Edwin Hewitt and Leonard J. Savage, *Symmetric measures on Cartesian products*, Trans. Amer. Math. Soc. **80** (1955), 470–501. MR0076206 (17,863g)

[46] Ulrich Hirth, *Exchangeable random ordered trees by positive definite functions*, J. Theoret. Probab. **16** (2003), no. 2, 339–344. MR1982030

[47] Ulrich Hirth and Paul Ressel, *Random partitions by semigroup methods*, Semigroup Forum **59** (1999), no. 1, 126–140. MR1847948

[48] ______, *Exchangeable random orders and almost uniform distributions*, J. Theoret. Probab. **13** (2000), no. 3, 609–634. MR1785522

[49] Saul Jacka and Jon Warren, *Random orderings of the integers and card shuffling*, Stochastic Process. Appl. **117** (2007), no. 6, 708–719. 2327835 (2008h:60022)

[50] Svante Janson, *Poset limits and exchangeable random posets*, Institut Mittag-Leffler preprint 02 (2009).

[51] Paul Joyce and Simon Tavaré, *A convergence theorem for symmetric functionals of random partitions*, J. Appl. Probab. **29** (1992), no. 2, 280–290. 1165214 (93g:60018)

[52] Olav Kallenberg, *Exchangeable random measures in the plane*, J. Theoret. Probab. **3** (1990), no. 1, 81–136. MR1031426

[53] ______, *Foundations of modern probability*, second ed., Probability and its Applications (New York), Springer-Verlag, New York, 2002. MR1876169

[54] ______, *Probabilistic symmetries and invariance principles*, Probability and its Applications (New York), Springer, New York, 2005. MR2161313

[55] S. V. Kerov, *Asymptotic representation theory of the symmetric group and its applications in analysis*, Translations of Mathematical Monographs, vol. 219, American Mathematical Society, Providence, RI, 2003, Translated from the Russian manuscript by N. V. Tsilevich, With a foreword by A. Vershik and comments by G. Olshanski. 1984868 (2005b:20021)

[56] J. F. C. Kingman, *The representation of partition structures*, J. London Math. Soc. (2) **18** (1978), no. 2, 374–380. MR509954

[57] ———, *The coalescent*, Stochastic Process. Appl. **13** (1982), no. 3, 235–248. MR671034

[58] ———, *Exchangeability and the evolution of large populations*, Exchangeability in probability and statistics (Rome, 1981), North-Holland, Amsterdam, 1982, pp. 97–112. MR675968

[59] E. Meeds, D. A. Ross, R. S. Zemel, and S. T. Roweis, *Learning stick-gure models using nonparametric bayesian priors over trees*, In IEEE Conference on Computer Vision and Pattern Recognition, 2008.

[60] Martin Möhle and Serik Sagitov, *A classification of coalescent processes for haploid exchangeable population models*, Ann. Probab. **29** (2001), no. 4, 1547–1562. MR1880231

[61] Robin Pemantle, *Automorphism invariant measures on trees*, Ann. Probab. **20** (1992), no. 3, 1549–1566. MR1175277

[62] J. Pitman, *Combinatorial stochastic processes*, Lecture Notes in Mathematics, vol. 1875, Springer-Verlag, Berlin, 2006, Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard. 2245368 (2008c:60001)

[63] Jim Pitman, *Exchangeable and partially exchangeable random partitions*, Probab. Theory Related Fields **102** (1995), no. 2, 145–158. 1337249 (96e:60059)

[64] Jim Pitman and Matthias Winkel, *Regenerative tree growth: binary self-similar continuum random trees and Poisson-Dirichlet compositions*, Ann. Probab. **37** (2009), no. 5, 1999–2041. MR2561439

[65] Paul Ressel, *Subdiagonal and almost uniform distributions*, Electron. Comm. Probab. **7** (2002), 97–101 (electronic). MR1902597

[66] ———, *Exchangeability and semigroups*, Rend. Mat. Appl. (7) **28** (2008), no. 1, 63–81. MR2463440

[67] Serik Sagitov, *The general coalescent with asynchronous mergers of ancestral lines*, J. Appl. Probab. **36** (1999), no. 4, 1116–1125. MR1742154

[68] Matthias Winkel and Bo Chen, *Restricted exchangeable partitions and embedding of associated hierarchies in continuum random trees*, On arxiv.org : arXiv:0911.5647v1.